GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2002, 04:52:54 ; Search time 63.01 Seconds
(without alignments)
824.989 Million cell updates/sec

Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MAMVFPSLVNGYDVAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2363	100.0	468	22	AAB66707	C.glutamicum phosp
2	2363	100.0	661	22	AAG92650	C glutamicum prote
3	2332	98.7	661	22	AAB69080	Brevibacterium lac
4	1821	77.1	362	22	AAB66708	C.glutamicum phosp
5	660	27.9	627	22	AAU37874	Streptococcus pneu
6	480.5	20.3	704	22	AAU60962	Propionibacterium
7	468.5	19.8	683	22	AAG93207	C glutamicum prote
8	468.5	19.8	683	22	AAB66721	C.glutamicum phosp
9	345.5	14.6	679	22	AAU34162	Staphylococcus aur
10	345.5	14.6	681	22	AAU36677	Staphylococcus aur
11	338.5	14.3	484	22	AAU35216	Enterococcus faeca
12	323	13.7	484	22	AAU36679	Staphylococcus aur
13	305.5	12.9	455	22	AAU38201	Salmonella typhi c
14	300	12.7	687	22	AAU34326	Staphylococcus aur
15	300	12.7	719	22	AAU37349	Staphylococcus aur
16	298.5	12.6	648	22	AAU34496	E. coli cellular p
17	298.5	12.6	648	22	AAG98282	Escherichia coli p
18	292	12.4	551	22	AAU34907	Enterococcus faeca
19	288.5	12.2	439	22	AAU34164	Staphylococcus aur
20	250.5	10.6	474	22	AAU34650	E. coli cellular p
21	247	10.5	523	22	AAU46075	Propionibacterium
22	234	9.9	135	22	AAB66723	C.glutamicum phosp
23	233.5	9.9	167	20	AAY20012	B. burgdorferi ant
24	232.5	9.8	583	21	AAY53620	Amino acid sequenc
25	229	9.7	280	22	AAU36886	Staphylococcus aur
26	228	9.6	205	22	AAU66872	Propionibacterium
27	228	9.6	207	22	AAU41453	Propionibacterium
28	226.5	9.6	842	22	AAU36357	Pseudomonas aerugi
29	221.5	9.4	263	22	AAU37425	Staphylococcus aur
30	221.5	9.4	263	22	AAU37559	Staphylococcus aur
31	220.5	9.3	135	22	AAU34346	Staphylococcus aur
32	201.5	8.5	142	22	AAU34222	Staphylococcus aur
33	197	8.3	168	22	AAU54924	Propionibacterium
34	189.5	8.0	173	22	ABG14977	Novel human diagno
35	177.5	7.5	285	22	ABG14978	Novel human diagno
36	168.5	7.1	118	22	AAU52635	Propionibacterium
37	166.5	7.0	135	20	AAY20013	B. burgdorferi ant
38	166.5	7.0	413	22	AAB66722	C.glutamicum phosp
39	165	7.0	640	22	AAU35193	Enterococcus faeca
40	161.5	6.8	126	18	AAW28051	Amino acid sequenc
41	156	6.6	449	21	AAY81556	Streptococcus pneu
42	154	6.5	108	21	AAB40791	Human ORFX ORF555
43	151.5	6.4	658	21	AAB15995	E. coli proliferat
44	149.5	6.3	375	22	ABG24413	Novel human diagno
45	141.5	6.0	639	22	ABB52613	Escherichia coli p

ALIGNMENTS

RESULT 1
AAB66707
ID AAB66707 standard; protein; 468 AA.
XX
AC AAB66707;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate protein #1.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-080989/09.
XX
PT Corynebacteium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation -
XX
PS Claim 4; Page 101-102; 144pp; English.
XX
CC The present invention relates to Corynebacteium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.
XX
SQ Sequence 468 AA;

Query Match 100.0%; Score 2363; DB 22; Length 468;
Best Local Similarity 100.0%; Pred. No. 3e-227;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mamvfpslvngydvaatmaagempmwslfgldvaqagyqgtvlpvlvvswilatiekflh 60

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 krlkgtadflitpvltllltgfltfiaigpamrwvgdvlahglqglydfggpvgglLfgl 120

Qy     121 VYSPIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 vyspivitglhqsfppielelfnqggsfifatasmaniaqgaaclavfflakseklkgla 180

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gasgvsavlgitepaifgvnlrlrwpffigigtaaiggalialfnikavalgaagflgvv 240

Qy     241 SIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEAP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 sidapdmvmflvcavvtffiafgaaiayglylvrrngsidpdataapvpagttkaeaeap 300

Qy     301 AEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 aefsndstiiqapltgeaialssvsdamfasgklgsgvaivptkgqlvspvsgkivvafp 360

Qy     361 SGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIK 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 sghafavrtkaedgsnvdilmhigfdtvnlngthfnplkkqgdevkagellcefdidaik 420

Qy     421 AAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPATP 468
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aagyevttpivvsnykktgpvntyglgeieaganllnvakkeavpatp 468
```

RESULT 2
AAG92650
ID AAG92650 standard; Protein; 661 AA.
XX
AC AAG92650;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6404.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH67869.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 661 AA;

Query Match 100.0%; Score 2363; DB 22; Length 661;
Best Local Similarity 100.0%; Pred. No. 5e-227;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     194 mamvfpslvngydvaatmaagempmwslfgldvaqagyqgtvlpvlvvswilatiekflh 253

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     254 krlkgtadflitpvltllltgfltfiaigpamrwvgdvlahglqglydfggpvgglLfgl 313

Qy     121 VYSPIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     314 vyspivitglhqsfppielelfnqggsfifatasmaniaqgaaclavfflakseklkgla 373

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     374 gasgvsavlgitepaifgvnlrlrwpffigigtaaiggalialfnikavalgaagflgvv 433

Qy     241 SIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEAP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     434 sidapdmvmflvcavvtffiafgaaiayglylvrrngsidpdataapvpagttkaeaeap 493

Qy     301 AEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     494 aefsndstiiqapltgeaialssvsdamfasgklgsgvaivptkgqlvspvsgkivvafp 553

Qy     361 SGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIK 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     554 sghafavrtkaedgsnvdilmhigfdtvnlngthfnplkkqgdevkagellcefdidaik 613

Qy     421 AAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPATP 468
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     614 aagyevttpivvsnykktgpvntyglgeieaganllnvakkeavpatp 661
```

```
RESULT   3
AAB69080
ID   AAB69080 standard; Protein; 661 AA.
XX
AC   AAB69080;
XX
DT   20-APR-2001  (first entry)
XX
DE   Brevibacterium lactofermentum sucrose PTS enzyme II protein SEQ ID NO:2.
XX
KW   Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrase;
KW   phosphoenolpyruvate:carbohydrate phosphotransferase system; glucose;
KW   coryneform bacterium; phosphoenolpyruvate-sugar transport system.
XX
OS   Brevibacterium lactofermentum.
XX
PN   WO200102584-A1.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-JP04348.
XX
PR   02-JUL-1999;   99JP-0189512.
XX
PA   (AJIN ) AJINOMOTO CO INC.
XX
PI   Izui M,  Sugimoto M,  Nakamatsu T,  Kurahashi O;
XX
DR   WPI; 2001-138150/14.
DR   N-PSDB; AAF32543.
XX
PT   Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT   II obtained by cassette ligation-mediated amplification of downstream
PT   domain of coryneform bacterium sucrase gene, with sucrose-binding
PT   activity -
XX
PS   Claim 1; Page 29-32; 45pp; Japanese.
XX
CC   The present sequence represents the Brevibacterium lactofermentum
CC   sucrose PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system
CC   or phosphoenolpyruvate-sugar transport system) enzyme II, which has
CC   sucrose-binding activity. A coryneform bacteria produced with the
CC   sucrose PTS enzyme II gene can have more efficient sugar uptake, and
CC   improved amino-acid and nucleic acid productivity. The sucrose PTS gene
CC   and it's disrupted gene, such as one without the sucrose PTS function,
CC   can be used to produce new breeds of coryneform bacterial strains to
CC   uptake sugar more efficiently e.g. glucose only or and sucrose, and can
CC   have improved amino-acid and nucleic acid productivity.
XX
SQ   Sequence   661 AA;


  Query Match             98.7%;  Score 2332;  DB 22;  Length 661;
  Best Local Similarity   98.5%;  Pred. No. 6.3e-224;
  Matches  461;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           ||||||:|||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     194 mamvfptlvngydvaatmtagempmwslfgldvaqagyqgtvlpvlvvswilatiekflh 253

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           ||| ||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db     254 krlmgtadflitpvltllltgfltfiaigpamrwvgdllahglqglydfggpvggllfgl 313

Qy     121 VYSPIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     314 vyspivitglhqsfppielelfnqggsfifatasmaniaqgaaclavfflakseklkgla 373

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
           ||||||||||||||||||||||||:||||||||||||:||||||||||||||||||||||
Db     374 gasgvsavlgitepaifgvnlrlrwpfyigigtaaiggalialfdikavalgaagflgvv 433

Qy     241 SIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEAP 300
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     434 sidapdmvmflvcavvtfviafgaaiayglylvrrngsidpdataapvpagttkaeaeap 493

Qy     301 AEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     494 aefsndstiiqapltgeaialssvsdamfasgklgsgvaivptkgqlvspvsgkivvafp 553

Qy     361 SGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIK 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     554 sghafavrtkaedgsnvdilmhigfdtvnlngthfnplkkqgdevkagellcefdidaik 613

Qy     421 AAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPATP 468
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     614 aagyevttpivvsnykktgpvntyglgeieaganllnvakkeavpatp 661


RESULT   4
AAB66708
ID   AAB66708 standard; protein; 362 AA.
XX
AC   AAB66708;
XX
DT   09-APR-2001  (first entry)
XX
DE   C.glutamicum phosphoenolpyruvate protein #2.
XX
KW   Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200102583-A2.
XX
PD   11-JAN-2001.
XX
PF   27-JUN-2000; 2000WO-IB00973.
XX
PR   01-JUL-1999;   99US-0142691.
PR   23-AUG-1999;   99US-0150310.
PR   03-SEP-1999;   99DE-1042095.
PR   03-SEP-1999;   99DE-1042097.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-080989/09.
XX
PT   Corynebacteium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT   sugar phosphotransferase system proteins or their portions, useful for
PT   typing or identifying C. glutamicum or related bacteria, and as markers
PT   for transformation -
XX
PS   Claim 6; Page 104-106; 144pp; English.
XX
CC   The present invention relates to Corynebacteium glutamicum
CC   phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC   The PTS nucleic acids and proteins are useful in the
CC   identification of microorganisms which can be used to produce fine
CC   chemicals, for modulating fine chemical production in C. glutamicum or
CC   related bacteria, the typing or identification of C. glutamicum or
CC   related bacteria, as reference points for mapping C. glutamicum genome,
CC   and as markers for transformation.
XX
SQ   Sequence   362 AA;


  Query Match             77.1%;  Score 1821;  DB 22;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 3e-173;
  Matches  362;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     107 YDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACLA 166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1 ydfggpvgqllfglvyspivitglhqsfppielelfnqggsfifatasmaniaqgaacla 60

Qy    167 VFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNI 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 vfflakseklkglagasgvsavlgitepaifgvnlrlrwpffigigtaaiggalialfni 120

Qy    227 KAVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAA 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 kavalgaagflgvvsidapdmvmflvcavvtffiafgaaiayglylvrrngsidpdataa 180

Qy    287 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 pvpagttkaeaeapaefsndstiiqapltgeaialssvsdamfasgklgsgvaivptkgq 240

Qy    347 LVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVK 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 lvspvsgkivvafpsghafavrtkaedgsnvdilmhigfdtvnlngthfnplkkqgdevk 300

Qy    407 AGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 agellcefdidaikaagyevttpivvsnykktgpvntyglgeieaganllnvakkeavpa 360

Qy    467 TP 468
          ||
Db    361 tp 362


RESULT   5
AAU37874
ID   AAU37874 standard; Protein; 627 AA.
XX
AC   AAU37874;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae cellular proliferation protein #303.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS55733.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 13467; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   627 AA;


  Query Match             27.9%;  Score 660;  DB 22;  Length 627;
  Best Local Similarity   34.3%;  Pred. No. 7.2e-57;
  Matches  159;  Conservative   92;  Mismatches  187;  Indels   26;  Gaps   10;

Qy      1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          | :|  || | : ||     ||:   : ||  :   | ||:||| ::  : |  || :
Db    182 mmlvsgslpnawava---qggevtamnffgf-ipvvglqgsvlpafiigvvgakfekavr 237

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          | :    | |:|| :|||:   |    |||    | : :    : :      :|| | |
Db    238 kvvpdvidllvtpfvtllvmsilglfvigpvfhvvenyiliatkailsmpfglggfligg 297

Qy    121 VYSPIVITGLHQSFPPIELELFNQGGSFIF-ATASMANIAQGAACLAVFFLAKSEKLKGL 179
          |:  ||::|:|  |  :|::|     :  | |  : |  ||||| :||    |: ||| |
Db    298 vhqlivvsgvhhifnllevqllaadhanpfnaiitaamtaqgaatvavgvktknpklktl 357

Qy    180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLG- 238
          |  : :|| ||||||||||||||| | |||: :   |||| | ::  :     |      |
Db    358 afpaalsaflgitepaifgvnlrfrkpfflsliagaiggglasilglagtgngitiipgt 417

Qy    239 VVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAE 298
          :: :    :  :|:   |:| : |     :|          : |||||   |   : : |
Db    418 mlyvgngqlpqyllmvavsfalgfaltymfgy-------edevdataaakqaevaeekee 470

Qy    299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVV 357
           |||   |:: :   |: |: :||: |:| :|:|| :| |:|: |::| : :    :: :
Db    471 vapaalqnetlv--tpivgdvvaladvndpvfssgamgqgiavkpsqgvvyaladaevsi 528

Qy    358 AFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDID 417
          |||:|||| ::|:    :  ::|:|:| |||::||  |     ||::||||::|  || :
Db    529 afptghafglktr----ngaevlihvgidtvsmngdgfeakvaqgnkvkagdvlgtfdsn 584

Qy    418 AIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV 458
          | ||| : || ::|:|   |    || |   | : |  :: |
Db    585 kiaaagldttmvivtntadyasvapvat---gsvskgdaviev 625


RESULT   6
AAU60962
ID   AAU60962 standard; Protein; 704 AA.
XX
AC   AAU60962;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #21858.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
```

PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59613.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 22157; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 704 AA;

Query Match 20.3%; Score 480.5; DB 22; Length 704;
Best Local Similarity 29.0%; Pred. No. 7.5e-39;
Matches 139; Conservative 80; Mismatches 209; Indels 51; Gaps 11;

```
Qy      17 TMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVLT 76
           |:   :: :  :||| :    |   | | |::: ||| : |:| | :      :  | ||
Db     242 tlggfkvnvvdvfglplтifnygsqvfpplimaailaplykwlkkiispnvqmifvpflt 301

Qy      77 LLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPP 136
           :|:   ||   |||   : |  ||:||: : |    :  :|  |:|  :|  |||
Db     302 mlimipltafligpigvyagaglanglkavndvspfifavliplpypfmvplglhwplna 361

Qy     137 IE-LELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEK----LKGLAGASGVSAVLGI 191
           :  | :   |  :|       | |   |   | ||:  :|     :   ||     : ||
Db     362 vmllniqtlgydyiqgpmgawnfacfgatagvlflsvrdkdvtmrqtaigalaagllggi 421

Qy     192 TEPAIFGVNLRLR--WPFFIGIGTAAIGGALIALFNIKAVALGAAG--------FLGVVS 241
           :||:::|::|| :  :|  : :|   :|| :|           || |        |  :::
Db     422 sepslygihlrykkiypsml-vg-clvggvii----------gaggglhtpafvftsllt 469

Qy     242 IDAPDMV-MFLVCAVVTFFIAFGAAIA---------YGLYLVRRNGSIDPDATAAPVPAG 291
           | | | : :: |    :|| |    |              ||  |        |  ||
Db     470 ipafdkiglysvaiaasfftamalvivrdyrdndekaaakarrrqasaheaeptaaapaa 529

Qy     292 TTK----------AEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIV 341
```

```
.          |:          |||   || : ::  : :|| | |: :| : | :|::| :| |:||
Db     530 atesitetatvegaeaptaaevaaeavtvtsplegravpiseipdpvfstgvvgdgiaie 589

Qy     342 PTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQ 401
           ||   :| |    :| |  ||||| ::      | |::|:|:| ||| | |  |:   |
Db     590 ptsntvvcpadatvvtapdsghafglkl----dsgvellihvgidtvelggkgfdvkvka 645

Qy     402 GDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           || | ||: |  ||   |  |||   ||::|:|      |      ::  |  :: | |
Db     646 gdhvsagqplvvfdptvideagyskitpvlvtngfdysnvtgipaddvttgtpvitvvk 704
```

RESULT 7
AAG93207
ID AAG93207 standard; Protein; 683 AA.
XX
AC AAG93207;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6961.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH68426.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 29; SEQ ID NO: 6961; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 683 AA;

Query Match 19.8%; Score 468.5; DB 22; Length 683;
Best Local Similarity 29.8%; Pred. No. 1.1e-37;
Matches 141; Conservative 79; Mismatches 206; Indels 47; Gaps 9;

```
Qy      27 SLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFI 86
           ::||| :    | | | | |: :  |  :|| | | :      :  |  :||:    |
Db     218 tvfglpmvlndysgqvfppliaaiglywvekglkkiipeavqmvfvpffsllimipataf 277

Qy      87 AIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQ-G 145
             :||    ||: ::: |: : :|   :  ::  |:|  :|  |||     | ::  |  |
Db     278 llgpfgigvgngisnlleainnfspfilsivipllypflvplglhwplnaimiqnintlg 337

Qy     146 GSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGAS--GVSAVL--GITEPAIFGVNL 201
             ||       | |       || |:  |: | :   |  |: | |  ||:||:::|| |
Db     338 ydfiqgpmgawnfacfglvtgvfllsikernkamrqvslggmlagllggisepslygvll 397

Qy     202 RLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIA 261
           | :  :|  :     || :: :|:||| |      | : ::|    : : :   | ||::
Db     398 rfkktyfrllpgclaggivmgifdikayafvftslltipamd--pwlgytigiavaffvs 455

Qy     262 FGAAIAYGLYLVRRNGSID-------------------PDAT-AAPV------------- 288
             :|      | |   |                      :|| ||||
Db     456 mflvlald---yrsneerdearakvaadkqaeedlkaeanatpaapvaaagagagagaga 512

Qy     289 PAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLV 348
            ||   | |  |   : :   | :|| |:|| || | | :||:|||| |:|| ||   :|
Db     513 aagaatavaakpklaagevvdivsplegkaiplsevpdpifaagklgpgiaiqptgntvv 572

Qy     349 SPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAG 408
           :|    :::   |||| |:|      | |:||:|:| ||| | |  |    ::  :||||
Db     573 apadatvilvqksghavalrl----dsgveilvhvgldtvqlggegftvhverrqqvkag 628

Qy     409 ELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKK 461
           : |  || | |::    : ||:||||  | | :      :  :   :: |  |
Db     629 dplitfdadfirskdlplitpvvvsnaakfgeiegipadqanssttvikvngk 681
```

RESULT 8
AAB66721
ID AAB66721 standard; protein; 683 AA.
XX
AC AAB66721;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate protein #5.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-080989/09.
XX
PT Corynebacteium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation -
XX
PS Claim 32; Page 136; 144pp; English.
XX
CC The present invention relates to Corynebacteium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.
XX
SQ Sequence 683 AA;

Query Match 19.8%; Score 468.5; DB 22; Length 683;
Best Local Similarity 29.8%; Pred. No. 1.1e-37;
Matches 141; Conservative 79; Mismatches 206; Indels 47; Gaps 9;

```
Qy      27 SLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFI 86
           ::||| :    | | | | |: :  |  :|| | | :      :  |  :||:    |
Db     218 tvfglpmvlndysgqvfppliaaiglywvekglkkiipeavqmvfvpffsllimipataf 277

Qy      87 AIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQ-G 145
             :||    ||: ::: |: : :|   :  ::  |:|  :|  |||     | ::  |  |
Db     278 llgpfgigvgngisnlleainnfspfilsivipllypflvplglhwplnaimiqnintlg 337

Qy     146 GSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGAS--GVSAVL--GITEPAIFGVNL 201
             ||       | |       || |:  |: | :   |  |: | |  ||:||:::|| |
Db     338 ydfiqgpmgawnfacfglvtgvfllsikernkamrqvslggmlagllggisepslygvll 397

Qy     202 RLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIA 261
           | :  :|  :     || :: :|:||| |      | : ::|    : : :   | ||::
Db     398 rfkktyfrllpgclaggivmgifdikayafvftsllti pamd--pwlgytigiavaffvs 455

Qy     262 FGAAIAYGLYLVRRNGSID-------------------PDAT-AAPV------------- 288
             :|      | |   |                      :|| ||||
Db     456 mflvlald---yrsneerdearakvaadkqaeedlkaeanatpaapvaaagagagagaga 512

Qy     289 PAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLV 348
            ||   | |  |   : :   | :|| |:|| || | | :||:|||| |:|| ||   :|
Db     513 aagaatavaakpklaagevvdivsplegkaiplsevpdpifaagklgpgiaiqptgntvv 572

Qy     349 SPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAG 408
           :|    :::   |||| |:|      | |:||:|:| ||| | |  |    ::  :||||
Db     573 apadatvilvqksghavalrl----dsgveilvhvgldtvqlggegftvhverrqqvkag 628

Qy     409 ELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKK 461
           : |  || | |::    : ||:||||  | | :      :  :   :: |  |
Db     629 dplitfdadfirskdlplitpvvvsnaakfgeiegipadqanssttvikvngk 681
```

RESULT 9
AAU34162
ID AAU34162 standard; Protein; 679 AA.
XX
AC AAU34162;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #438.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.

```
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52021.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5658; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   679 AA;


  Query Match             14.6%;  Score 345.5;  DB 22;  Length 679;
  Best Local Similarity   24.2%;  Pred. No. 2.2e-25;
  Matches  132;  Conservative   68;  Mismatches  192;  Indels  153;  Gaps   17;

Qy      27 SLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL----ITPVLTLLLTGF 82
           |: |:   | |  | ::   : :|     || :  |     |       |:: :  | |
Db     123 silgiptlqtgvfggiiigalaaw---cynkfyninlpsylgffagkrfvpim-mattsf 178

Qy      83 LTFIAIGPAMRWVGDVLAHGLQ----GLYDFGGPVGGLLFGLVYSPIVITGLHQSF---- 134
           :  :|   |: |    :  ||     || |    |   ||| :   ::  |||  |
Db     179 i--lafpmaliw--ptiqtglnafstglldsntgvavflfgfikrllipfglhhifhapf 234

Qy     135 ------------------PPIELELFNQG-----GSFIFATASMANIAQGAACLAVFFLA 171
                            | :|   :|     | |:     :      || ||::  |
Db     235 wfefgswknaageiihgdqrifieqiregahltagkfmqgefpvmmfglpaaalaiyhsa 294

Qy     172 KSEKLKGLAGASGVSA----VLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIK 227
           | |  | :||  | :|    : |||||  |         |||    | : |       :
Db     295 kpenkkvvaglmgsaaltsfltgiteplefsflfvapllffi---havldglsfltlyll 351

Qy     228 AVALG---AAGFLGVVSIDA-PDMVMF-------LVCAVVTFFI---------------- 260
           | ||    : ||:  | :   |:   :        || ||: :|:
Db     352 dvhlgytfsggfidyvllgvlpnktqwwlvipvglvyaviyyfvfrflivklkyktpgre 411

Qy     261 --------AFGAAIAYGLY---------------LVR---------------------- 274
                   |    : | :               : |
Db     412 dkqsqavtasatelpyavleamggkanikhldacitrlrvevndkskvdvpglkdlgasg 471

Qy     275 -------------------------RNGSIDPDATAAPVPAGTTKAEAEAPAEFSNDST 308
                                    || :  : |       |   ||  :  |  |
Db     472 vlevgnnmqaifgpksdqikhemqqimngqvvenpttmeddkdetvvvaedksatselsh 531

Qy     309 IIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVR 368
           |: ||||||   || | | :|:   :| |:|| |::|:: :| :||: : ||: || .:

Db     532 ivhapltgevtplsevpdqvfsekmmgdgiaikpsqgevrapfngkvqmifptkhaigl- 590

Qy     369 TKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIKAAGYEVTT 428
                 | :::|:||| ||| |||  |    ::| ||| |:|| ||:| |:       |
Db     591 ---vsdsglellihigldtvklngegftlhveegqevkqgdllinfdldyirnhaksdit 647

Qy     429 PIVVS 433
              ||:|:
Db     648 piivt 652


RESULT  10
AAU36677
ID   AAU36677 standard; Protein; 681 AA.
XX
AC   AAU36677;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #847.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54536.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12270; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence   681 AA;

  Query Match             14.6%;  Score 345.5;  DB 22;  Length 681;
  Best Local Similarity   24.2%;  Pred. No. 2.2e-25;
  Matches  132;  Conservative   68;  Mismatches  192;  Indels  153;  Gaps   17;

Qy      27 SLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL----ITPVLTLLLTGF 82
           |: |:   | |  | ::   : :|      || :  |      |       |:: :  | |
Db     125 silgiptlqtgvfggiiigalaaw---cynkfyninlpsylgffagkrfvpim-mattsf 180

Qy      83 LTFIAIGPAMRWVGDVLAHGLQ----GLYDFGGPVGGLLFGLVYSPIVITGLHQSF---- 134
           :  :|   |: |    :  ||     || |    |   ||| :   ::  |||  |
Db     181 i--lafpmaliw--ptiqtglnafstglldsntgvavflfgfikrllipfglhhifhapf 236

Qy     135 ------------------PPIELELFNQG-----GSFIFATASMANIAQGAACLAVFFLA 171
                             | :|   :|     | |:     :      || ||::  |
Db     237 wfefgswknaageiihgdqrifieqiregahltagkfmqgefpvmmfglpaaalaiyhsa 296

Qy     172 KSEKLKGLAGASGVSA----VLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIK 227
           | |  | :||  | :|    : |||||  |          |||    | : |       :
Db     297 kpenkkvvaglmgsaaltsfltgiteplefsflfvapllffi---havldglsfltlyll 353

Qy     228 AVALG---AAGFLGVVSIDA-PDMVMF-------LVCAVVTFFI---------------- 260
           | ||   : ||:  | :   |:   :        || ||: :|:
Db     354 dvhlgytfsggfidyvllgvlpnktqwwlvipvglvyaviyyfvfrflivklkyktpgre 413

Qy     261 --------AFGAAIAYGLY---------------LVR---------------------- 274
                   |    : | :                : |
Db     414 dkqsqavtasatelpyavleamggkanikhldacitrlrvevndkskvdvpglkdlgasg 473

Qy     275 -------------------------RNGSIDPDATAAPVPAGTTKAEAEAPAEFSNDST 308
                               || :  : |        |   ||  :  |  |
Db     474 vlevgnnmqaifgpksdqikhemqqimngqvvenpttmeddkdetvvvaedksatselsh 533

Qy     309 IIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVR 368
           |: ||||||   || | | :|:   :| |:|| |::|:: :| :||: : ||: ||  :
Db     534 ivhapltgevtplsevpdqvfsekmmgdgiaikpsqgevrapfngkvqmifptkhaigl- 592

Qy     369 TKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIKAAGYEVTT 428
               | :::|:||| ||| |||  |    ::| ||| |:||  ||:| |:       |
Db     593 ---vsdsglellihigldtvklngegftlhveegqevkqgdllinfdldyirnhaksdit 649

Qy     429 PIVVS 433
           ||:|:
Db     650 piivt 654


RESULT  11
AAU35216
ID   AAU35216 standard; Protein; 484 AA.
XX
AC   AAU35216;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Enterococcus faecalis cellular proliferation protein #503.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Enterococcus faecalis.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS53075.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 10809; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   484 AA;

  Query Match             14.3%;  Score 338.5;  DB 22;  Length 484;
  Best Local Similarity   33.2%;  Pred. No. 6.6e-25;
  Matches   94;  Conservative   54;  Mismatches  114;  Indels   21;  Gaps   10;

Qy       2 AMVF---PSLVNGYDVAATMAAG---EMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATI 55
           | ||   |:|  |   | ||  |   : |: ::|      || || :: |:   |:|: :
Db     201 asvfgatpal-ggvigavtmltgmnpdapisniftggtlsag-qggiigvifavwllsll 258

Qy      56 EKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDF----GG 111
           || ||| :  : | ::|| ::||: |  |   | |    |   :::|| |: :     ||
Db     259 ekqlhkivpesidiivtptisllviglatiflimp----vagaisnglvgiinvvlekgg 314

Qy     112 PVGGLLFGLVYSPIVITGLHQSFPPIELELFNQGG-SFIFATASMANIAQGAACLAVFFL 170
             | |   || : |:|: ||||   || :|: || | : :    :||   |  | ||::
Db     315 mvagftlgltflpmvmfglhqiltpihieminqtgmtlllpilamagagqvgaalalwir 374

Qy     171 AKSEK--LKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALF-NIK 227
             ||:|  :: : ||  |  :||| || |:|| | |  ||         ||||:|  | |:
Db     375 cksdkklvemikgalpv-gilgigepliygvtlplgrpfitacigggiggavigafgnvg 433

Qy     228 AVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGL 270
           |:|:| :|   :  |     : :::  :  :   | | : :|:
Db     434 aiaigpsgvalipliannqwlayvlgllaayaggfvatlffgi 476


RESULT  12
AAU36679
ID   AAU36679 standard; Protein; 484 AA.
XX
AC   AAU36679;
XX
DT   14-FEB-2002  (first entry)
XX
```

```
DE   Staphylococcus aureus cellular proliferation protein #849.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54538.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12272; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   484 AA;


  Query Match           13.7%;  Score 323;  DB 22;  Length 484;
  Best Local Similarity  32.2%;  Pred. No. 2.4e-23;
  Matches   96;  Conservative   45;  Mismatches  115;  Indels   42;  Gaps   10;

Qy       9 VNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTAD 68
Db     209 viggttlltgiagknilmnvftgeplqpg-qggiigvifavwilsivekrlhkivpnaid 267

Qy      69 FLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVIT 128
Db     268 iivtptiallivglltififmplagfvsdslvsvvngiisiggvfsgfiigasflplvml 327

Qy     129 GLHQSFPPIELELFNQ-GGSFIFATASMANIAQGAACLAVFFLAKSEKL--KGLAGASGV 185
Db     328 glhhiftpihieminqsgatyllpiaamagagqvgaalalwvrckrnttlrntlkgalpv 387

Qy     186 SAVLGITEPAIFGVNLRLRWPFFI-----GIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
Db     388 -gflgigepliygvtlplgrpfltacigggiggavigg----ighigakaigpsg-vsll 441

Qy     241 SIDAPDMVMFLVCAVVTFFIAFGAAIAYGL---YLVRRNGSIDPDATAAPVPAGTTKA 295
Db     442 plisdnmylgyiagll-------aayaggfvctylf----------------gttka 475


RESULT  13
AAU38201
ID   AAU38201 standard; Protein; 455 AA.
XX
AC   AAU38201;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Salmonella typhi cellular proliferation protein #92.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Salmonella typhi.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS56060.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 13794; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   455 AA;
```

```
  Query Match             12.9%;  Score 305.5;  DB 22;  Length 455;
  Best Local Similarity   31.3%;  Pred. No. 1.2e-21;
  Matches   79;  Conservative   55;  Mismatches  111;  Indels    7;  Gaps    5;

Qy      14 VAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITP 73
           :|  ::: ::   :||| :  | | :| |: ||:|  ::  ||:   | | |: : :: |
Db     194 magilsspqlaqitlfg-ealqpg-rggviavllvvalmcwierqfrkllpgslelilnp 251

Qy      74 VLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQS 133
           :|| ::|| :  :|: |   |: | :|||     | || : | :    : |:|: ||||
Db     252 llttvitgavaivalqplggwisdaiahgaswaidrggflvgavlagtflplvliglhqg 311

Qy     134 FPPIELELFN-QGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGIT 192
             | :||      : :|   :|| : |  | :||    :: :|| :   :    :|||
Db     312 lvxihvelvqahxynalfpilamagvgqigaaiavlmktrnarlkkvikgalpvgllgig 371

Qy     193 EPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAV-ALGAAGFLGVVSIDAPDMVMFL 251
           || |||| | |  ||       |:|||||: : :  |   | :|    ::| |  :: :|
Db     372 eplifgvtlplgkpfigaclggavggalisywkvatvitfgisglplaltivagkvlfyl 431

Qy     252 ---VCAVVTFFI 260
               : ||:  ||
Db     432 lgyliaviagfi 443


RESULT  14
AAU34326
ID   AAU34326 standard; Protein; 687 AA.
XX
AC   AAU34326;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #602.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52185.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5822; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   687 AA;


  Query Match             12.7%;  Score 300;  DB 22;  Length 687;
  Best Local Similarity   22.0%;  Pred. No. 7.9e-21;
  Matches  124;  Conservative   88;  Mismatches  181;  Indels  170;  Gaps   21;

Qy      17 TMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL----IT 72
           : || |     : |:   | |  | ::   : :|      || :  |     |
Db     127 sqaakepahalvlgiptlqtgvfggiimgalaaw---cynkfynitlppflgffagkrfv 183

Qy      73 PVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDF----GGPVGGLLFGLVYSPIVIT 128
           |::| ::      || |  : :    :  ||  | :|      :   :||::   ::
Db     184 pivtsvva-----iatgvllsfawppiqdglnslsnfllnknltlttfifgiierslipf 238

Qy     129 GLHQSF-PPIELEL---FNQGGSFIFATAS--MANIAQG---------------------- 161
           |||  |  |   |     |  |  :       || :  |
Db     239 glhhifyspfwfefgsytnhagelvrgdqriwmaqlkdgvpftagafttgkypfmmfglp 298

Qy     162 AACLAVFFLAKSEKLK---GLAGASGVSAVL-GITE----------PAIFGVNLRLRWPF 207
           ||  |::  |: |: |   ||  ::|::| | ||||          | ::|::: |
Db     299 aaafaiyknarperkkvvgglmlsagltafltgiteplefsflfvapvlygihvllagts 358

Qy     208 FIGIGTAAIGGALIALFNIKAVALGAAGFL-----GVVSIDAPDMVMF----LVCAVVTF 258
           |:          :: |  :|     : ||:     |::: |    ::     :| |:| :
Db     359 fl----------vmhllgvkigmtfsggfidyilyglllnwdrshallvipvgivyaivyy 408

Qy     259 FIAFGAAIAYGLYLVRRNGSIDPD-----ATAAPVP---------------------- 289
           |: |  ||    : ::  |  | :     :: | :|
Db     409 fl-fdfair--kfklktpgredeeteirnssvaklpfdvldamggkenikhldacitrlr 465

Qy     290 -----------AG-------------------------------------TTKAEAEA 299
                      ||                                     |  :|
Db     466 vevvdkskvdvagikalgasgvlevgnnmqaifgpksdqikhdmakimsgeitkpsettv 525

Qy     300 PAEFSND--------STIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPV 351
             | |::        :| | ||  |: | || | | :||   :| ||  :| ||::|:|
Db     526 teemsdepvhvealgttdiyapgigqiiplsevpdqvfagkmmgdgvgfipekgeivapf 585

Qy     352 SGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELL 411
           | :   ||: ||  :   |  | |::|:||| ||| |||  |  |    ::|  |: |
Db     586 dgtvktifptkhaigl----esesgvevlihigidtvklngegfeslinvdekvtqgqpl 641

Qy     412 CEFDIDAIKAAGYEVTTPIVVSN 434
           : ::  :||    : ||::::|
Db     642 mkvnlaylkahapsivtpmiitn 664


RESULT  15
AAU37349
ID   AAU37349 standard; Protein; 719 AA.
XX
AC   AAU37349;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1519.
XX
```

```
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS55208.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12942;. 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   719 AA;
```

```
  Query Match           12.7%;  Score 300;  DB 22;  Length 719;
  Best Local Similarity  22.0%;  Pred. No. 8.4e-21;
  Matches  124;  Conservative   88;  Mismatches  181;  Indels  170;  Gaps   21;

Qy      17 TMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL----IT 72
           : || |     : |:   | |  | ::   : :|      || :  |     |
Db     158 sqaakepahalvlgiptlqtgvfggiimgalaaw---cynkfynitlppflgffagkrfv 214

Qy      73 PVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDF----GGPVGGLLFGLVYSPIVIT 128
           |::| ::      || |  : :    :  ||  | :|       :   :||::   ::
Db     215 pivtsvva-----iatgvllsfawppiqdglnslsnfllnknltlttfifgiierslipf 269

Qy     129 GLHQSF-PPIELEL---FNQGGSFIFATAS--MANIAQG---------------------- 161
           |||  |  |   |     |  |  :      || :  |
Db     270 glhhifyspfwfefgsytnhagelvrgdqriwmaqlkdgvpftagafttgkypfmmfglp 329

Qy     162 AACLAVFFLAKSEKLK---GLAGASGVSAVL-GITE----------PAIFGVNLRLRWPF 207
           ||  |::  |: |: |   ||  ::|::| | ||||          | ::|::: |

Db     330 aaafaiyknarperkkvvgglmlsagltafltgiteplefsflfvapvlygihvllagts 389

Qy     208 FIGIGTAAIGGALIALFNIKAVALGAAGFL-----GVVSIDAPDMVMF----LVCAVVTF 258
           |:          :: |  :|     : ||:     |::: |    ::     :| |:| :
Db     390 fl----------vmhllgvkigmtfsggfidyilygllnwdrshallvipvgivyaivyy 439

Qy     259 FIAFGAAIAYGLYLVRRNGSIDPD-----ATAAPVP---------------------- 289
           |: |  ||    : ::  |  | :     :: | :|
Db     440 fl-fdfair--kfklktpgredeeteirnssvaklpfdvldamggkenikhldacitrlr 496

Qy     290 -----------AG---------------------------------------TTKAEAEA 299
                      ||                                       |  :|
Db     497 vevvdkskvdvagikalgasgvlevgnnmqaifgpksdqikhdmakimsgeitkpsettv 556

Qy     300 PAEFSND--------STIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPV 351
             | |::         :| | ||  |: | || | | :||   :| ||  :| ||::|:|
Db     557 teemsdepvhvealgttdiyapgigqiiplsevpdqvfagkmmgdgvgfipekgeivapf 616

Qy     352 SGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELL 411
            | :   ||: ||  :    |  | |::|:||| ||| |||  |  |    ::|  |: |
Db     617 dgtvktifptkhaigl----esesgvevlihigidtvklngegfeslinvdekvtqgqpl 672

Qy     412 CEFDIDAIKAAGYEVTTPIVVSN 434
            : ::  :||    : ||::::|
Db     673 mkvnlaylkahapsivtpmiitn 695
```

Search completed: September 25, 2002, 06:00:08
Job time: 4034 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2002, 05:22:04 ; Search time 24.86 Seconds
(without alignments)
459.822 Million cell updates/sec

Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MAMVFPSLVNGYDVAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	113.5	4.8	1176	1	US-07-828-788A-12	Sequence 12, Appl
2	113.5	4.8	1176	1	US-08-356-034-2	Sequence 2, Appli
3	113.5	4.8	1176	3	US-08-933-891-2	Sequence 2, Appli
4	113.5	4.8	1176	5	PCT-US92-11337-12	Sequence 12, Appl
5	113.5	4.8	1179	1	US-08-040-751-1	Sequence 1, Appli
6	108.5	4.6	731	2	US-08-911-364-1	Sequence 1, Appli
7	108.5	4.6	733	4	US-08-464-700-2	Sequence 2, Appli
8	108	4.6	1011	3	US-08-836-325-2	Sequence 2, Appli
9	108	4.6	1989	3	US-08-836-325-12	Sequence 12, Appl
10	107	4.5	1984	3	US-08-836-325-10	Sequence 10, Appl
11	106.5	4.5	1129	6	5164180-6	Patent No. 5164180
12	106.5	4.5	1179	6	5188960-2	Patent No. 5188960
13	106.5	4.5	1864	2	US-08-804-227C-3	Sequence 3, Appli
14	105.5	4.5	462	3	US-08-801-344-9	Sequence 9, Appli
15	105.5	4.5	462	4	US-09-498-599-9	Sequence 9, Appli
16	103.5	4.4	396	2	US-08-850-880-4	Sequence 4, Appli
17	103.5	4.4	396	2	US-08-944-916-4	Sequence 4, Appli
18	103.5	4.4	396	2	US-08-814-877-4	Sequence 4, Appli
19	102	4.3	3033	1	US-07-925-695-9	Sequence 9, Appli
20	101	4.3	3033	1	US-07-925-695-8	Sequence 8, Appli
21	100.5	4.3	412	4	US-09-445-472-1	Sequence 1, Appli
22	100.5	4.3	522	4	US-08-894-818B-3	Sequence 3, Appli
23	100.5	4.3	522	4	US-09-445-472-4	Sequence 4, Appli
24	100.5	4.3	525	1	US-08-356-340-2	Sequence 2, Appli
25	100.5	4.3	525	2	US-08-786-555-2	Sequence 2, Appli
26	100.5	4.3	654	4	US-08-894-818B-35	Sequence 35, Appl
27	100.5	4.3	654	4	US-09-445-472-16	Sequence 16, Appl
28	100.5	4.3	713	3	US-09-335-409-11	Sequence 11, Appl
29	100.5	4.3	713	4	US-09-568-102-11	Sequence 11, Appl
30	100.5	4.3	713	4	US-09-567-969-11	Sequence 11, Appl
31	100.5	4.3	713	4	US-09-568-480-11	Sequence 11, Appl
32	100.5	4.3	713	4	US-09-568-486-11	Sequence 11, Appl
33	100.5	4.3	713	4	US-09-568-472-11	Sequence 11, Appl
34	100	4.2	528	2	US-08-403-852D-21	Sequence 21, Appl
35	100	4.2	528	3	US-08-510-646B-22	Sequence 22, Appl
36	100	4.2	528	4	US-09-231-818-21	Sequence 21, Appl
37	100	4.2	1969	3	US-08-836-325-16	Sequence 16, Appl
38	99.5	4.2	792	2	US-08-678-039A-40	Sequence 40, Appl
39	99	4.2	756	4	US-08-960-048-10	Sequence 10, Appl
40	99	4.2	1275	3	US-09-120-513-2	Sequence 2, Appli
41	99	4.2	1275	4	US-09-450-105-2	Sequence 2, Appli
42	98	4.1	1184	6	5254799-6	Patent No. 5254799
43	98	4.1	1188	6	5254799-7	Patent No. 5254799
44	97.5	4.1	2005	3	US-08-836-325-7	Sequence 7, Appli
45	96.5	4.1	593	6	5523211-1	Patent No. 5523211

ALIGNMENTS

```
RESULT 1
US-07-828-788A-12
; Sequence 12, Application US/07828788A
; Patent No. 5273746
;  GENERAL INFORMATION:
;    APPLICANT:  PAYNE, JEWEL M.
;    APPLICANT:  HICKLE, LESLIE A.
;    TITLE OF INVENTION:  NOVEL BACILLUS THURINGIENSIS ISOLATES
;    TITLE OF INVENTION:  ACTIVE AGAINST PHTHIRAPTERA PESTS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID R. SALIWANCHIK
;      STREET:  2421 N.W. 41st STREET, SUITE A-1
;      CITY:  GAINESVILLE
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/828,788A
;      FILING DATE:  19920129
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SALIWANCHIK, DAVID R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  MA75
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1176 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  BACILLUS THURINGIENSIS
;      STRAIN:  AIZAWAI
;      INDIVIDUAL ISOLATE:  PS81I
;    IMMEDIATE SOURCE:
;      LIBRARY:  LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK
;      CLONE:  81IA2
US-07-828-788A-12
```

```
  Query Match             4.8%;  Score 113.5;  DB 1;  Length 1176;
  Best Local Similarity   22.4%;  Pred. No. 0.021;
  Matches   56;  Conservative   43;  Mismatches   96;  Indels   55;  Gaps   13;

Qy     249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
           :|::      | :  | :   :|  |: |::|      |   : |  ||        |
Db     377 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 434

Qy     295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
           ::|        ||        |||||    | | | ||| ::| | | :  :   |  |
Db     435 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 493

Qy     337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
           :  : : ||: :    :: :  |    : || :    |  ::  |   |   :|  :|:
Db     494 DILRITSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 548

Qy     397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
             |  :::|         : : ||:  |||    |:     : |       :| | :
Db     549 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 593

Qy     457 NVAKKEAVPA 466
           : : | |||
Db     594 YIERIEFVPA 603


RESULT   2
US-08-356-034-2
; Sequence 2, Application US/08356034
; Patent No. 5691308
;  GENERAL INFORMATION:
;    APPLICANT:  Payne, Jewel M.
;    APPLICANT:  Sick, August J.
;    TITLE OF INVENTION:  No. 5691308el Bacillus thuringiensis Isolate
;    TITLE OF INVENTION:  Active Against Lepidopteran Pests, and Genes Encoding
;    TITLE OF INVENTION:  No. 5691308el Lepidopteran-Active Toxins
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  US
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,034
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/210,110
;      FILING DATE:
;      APPLICATION NUMBER:  07/865,168
;      FILING DATE:  09-APR-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/451,261
;      FILING DATE:  14-DEC-89
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/371,955
;      FILING DATE:  27-JUN-89
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, Roman
;      REGISTRATION NUMBER:  21,023
;      REFERENCE/DOCKET NUMBER:  MA43.C1.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904)375-8100
;      TELEFAX:  (904)372-5800
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1176 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  BACILLUS THURINGIENSIS
;      STRAIN:  AIZAWAI
;      INDIVIDUAL ISOLATE:  PS81I
;    IMMEDIATE SOURCE:
;      LIBRARY:  LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK
;      CLONE:  81IA2
US-08-356-034-2


  Query Match             4.8%;  Score 113.5;  DB 1;  Length 1176;
  Best Local Similarity   22.4%;  Pred. No. 0.021;
  Matches   56;  Conservative   43;  Mismatches   96;  Indels   55;  Gaps   13;

Qy     249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
           :|::      | :  | :   :|  |: |::|      |   : |  ||        |
Db     377 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 434

Qy     295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
           ::|        ||        |||||    | | | ||| ::| | | :  :   |  |
Db     435 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 493

Qy     337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
           :  : : ||: :    :: :  |    : || :    |  ::  |   |   :|  :|:
Db     494 DILRITSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 548

Qy     397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
             |  :::|         : : ||:  |||    |:     : |       :| | :
Db     549 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 593

Qy     457 NVAKKEAVPA 466
           : : | |||
Db     594 YIERIEFVPA 603


RESULT   3
US-08-933-891-2
; Sequence 2, Application US/08933891
; Patent No. 6096708
;  GENERAL INFORMATION:
;    APPLICANT:  Payne, Jewel M.
;    APPLICANT:  Sick, August J.
;    TITLE OF INVENTION:  No. 6096708el Bacillus thuringiensis Isolate
;    TITLE OF INVENTION:  Active Against Lepidopteran Pests, and Genes Encoding
;    TITLE OF INVENTION:  No. 6096708el Lepidopteran-Active Toxins
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  US
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,891
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/356,034
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/210,110
;      FILING DATE:
;      APPLICATION NUMBER:  07/865,168
;      FILING DATE:  09-APR-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/451,261
;      FILING DATE:  14-DEC-89
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/371,955
;      FILING DATE:  27-JUN-89
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, Roman
;      REGISTRATION NUMBER:  21,023
;      REFERENCE/DOCKET NUMBER:  MA43.C1.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904)375-8100
;      TELEFAX:  (904)372-5800
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1176 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  BACILLUS THURINGIENSIS
;      STRAIN:  AIZAWAI
;      INDIVIDUAL ISOLATE:  PS81I
;    IMMEDIATE SOURCE:
;      LIBRARY:  LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK
;      CLONE:  81IA2
US-08-933-891-2


  Query Match             4.8%;  Score 113.5;  DB 3;  Length 1176;
  Best Local Similarity   22.4%;  Pred. No. 0.021;
  Matches   56;  Conservative   43;  Mismatches   96;  Indels   55;  Gaps   13;

Qy     249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
           :|::      | :  | :   :|  |: |::|      |   : |  ||        |
Db     377 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 434

Qy     295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
           ::|        ||        |||||    | | | ||| ::| | | :  :   |  |
Db     435 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 493

Qy     337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
           :  : : ||: :    :: :  |    : || :    |  ::  |   |   :|  :|:
Db     494 DILRITSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 548

Qy     397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
             |  :::|         : : ||:  |||    |:     : |       :| | :
Db     549 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 593

Qy     457 NVAKKEAVPA 466
           : : | |||
Db     594 YIERIEFVPA 603


RESULT   4
PCT-US92-11337-12
; Sequence 12, Application PC/TUS9211337
;  GENERAL INFORMATION:
;    APPLICANT:  PAYNE, JEWEL M.
;    APPLICANT:  HICKLE, LESLIE A.
;    TITLE OF INVENTION:  NOVEL BACILLUS THURINGIENSIS ISOLATES
;    TITLE OF INVENTION:  ACTIVE AGAINST PHTHIRAPTERA PESTS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID R. SALIWANCHIK
;      STREET:  2421 N.W. 41st STREET, SUITE A-1
;      CITY:  GAINESVILLE
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/11337
;      FILING DATE:  19921231
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  97/828,788
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SALIWANCHIK, DAVID R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  MA75
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1176 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  BACILLUS THURINGIENSIS
;      STRAIN:  AIZAWAI
;      INDIVIDUAL ISOLATE:  PS81I
;    IMMEDIATE SOURCE:
;      LIBRARY:  LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK
;      CLONE:  81IA2
PCT-US92-11337-12


  Query Match             4.8%;  Score 113.5;  DB 5;  Length 1176;
  Best Local Similarity   22.4%;  Pred. No. 0.021;
  Matches   56;  Conservative   43;  Mismatches   96;  Indels   55;  Gaps   13;

Qy     249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
           :|::      | :  | :   :|  |: |::|      |   : |  ||        |
Db     377 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 434

Qy     295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
           ::|        ||        |||||    | | | ||| ::| | | :  :   |  |
Db     435 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 493

Qy     337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
           :  : : ||: :    :: :  |    : || :    |  ::  |   |   :|  :|:
Db     494 DILRITSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 548

Qy     397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
             |  :::|         : : ||:  |||    |:     : |       :| | :
Db     549 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 593

Qy     457 NVAKKEAVPA 466
           : : | |||
Db     594 YIERIEFVPA 603


RESULT   5
US-08-040-751-1
; Sequence 1, Application US/08040751
; Patent No. 5407825
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Payne, Jewel M
;    APPLICANT:  Sick, August J
;    TITLE OF INVENTION:  No. 5407825el Bacillus thuringiensis isolates
;    TITLE OF INVENTION:  active against Lepidopteran Pests and Genes Encoding No. 54078
;    TITLE OF INVENTION:  Lepidopteran-active Toxins
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID R. SALIWANCHIK
;      STREET:  2421 N.W. 41st STREET, SUITE A-1
;      CITY:  GAINESVILLE
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/040,751
;      FILING DATE:  19930329
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SALIWANCHIK, DAVID R.
;      REFERENCE/DOCKET NUMBER:  MA39.C1.D3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1179 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  BACILLUS THURINGIENSIS
;      STRAIN:  AIZAWAI
;      INDIVIDUAL ISOLATE:  PS81I
;    IMMEDIATE SOURCE:
;      LIBRARY:  LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK
;      CLONE:  81RR1
US-08-040-751-1


  Query Match             4.8%;  Score 113.5;  DB 1;  Length 1179;
  Best Local Similarity   22.4%;  Pred. No. 0.021;
  Matches   56;  Conservative   43;  Mismatches   96;  Indels   55;  Gaps   13;

Qy    249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
          :|::      | :  | :   :|  |: |::|      |   : |  ||         |
Db    380 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 437

Qy    295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
          ::|       ||        |||||    | | | ||| ::| | | :  :   |  |
Db    438 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 496

Qy    337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
           :  : : ||: :    :: :  |    : || :    |  ::  |   |   :|  :|:
Db    497 DILRITSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551

Qy    397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
              |  :::|         : : ||:  |||    |:      : |        :| | :
Db    552 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 596

Qy    457 NVAKKEAVPA 466
           : : | |||
Db    597 YIERIEFVPA 606
```

```
RESULT   6
US-08-911-364-1
; Sequence 1, Application US/08911364
; Patent No. 5969106
;  GENERAL INFORMATION:
;    APPLICANT:  ROTHSTEIN, Aser
;    APPLICANT:  KEELY, Fred W.
;    APPLICANT:  ROTHSTEIN, Steven J.
;    TITLE OF INVENTION:  SELF-ALIGNING PEPTIDES MODELED ON HUMAN
;    TITLE OF INVENTION:  ELASTIN AND OTHER FIBROUS PROTEINS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/911,364
;      FILING DATE:  07-AUG-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/023,552
;      FILING DATE:  07-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  041082/0104
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  731 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-911-364-1


  Query Match             4.6%;  Score 108.5;  DB 2;  Length 731;
  Best Local Similarity   20.6%;  Pred. No. 0.032;
  Matches  122;  Conservative   47;  Mismatches  227;  Indels  195;  Gaps   22;

Qy     16 ATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVL 75
          | :  |::|   | |:      | | |||                  |  |
Db    108 AGVKPGKVPGVGLPGV------YPGGVLP--------------GARFPGVG-------- 138

Qy     76 TLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL-VYSPIVITGLHQSF 134
            :| |  |   : |    ||   | |: |:  ||||  |:  |  : :| :  |    :
Db    139 --VLPGVPTGAGVKPKAPGVGGAFA-GIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPY 195

Qy    135 PPIELELFNQGGGSFIFATASMANIAQG-----------AACLAVFFLAKSEKLKGLAGA 182
             :|  :  |   :   |  |    |             |   | |    :  | |: ||
Db    196 TTGKLP-YGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAAAKAAAKFGAGAAGVLPGVGGA 254

Qy    183 SGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFL----- 237
           ||  | |   | | |:          |:|| |   |  |   ||   |||  |
Db    255 -GVPGVPGAI-PGIGGI---------AGVGTPAAAAAAAAA--AKAAKYGAAAGLVPGGP 301

Qy    238 ------------------------------------GVVSIDAPDMVMFL-------- 251
                                               |||| :|
Db    302 GFGPGVVGVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARP 361

Qy    252 ---VCAVVTFFIAFGAAIAYGLYLVRRNG-------------------SIDPDATAAP-- 287
```

```
Db    362 GVGVGGIPTYGVGAGGFPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAA 421

Qy    288 --------VPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASG-KLGSGV 338
Db    422 KAAKYGVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGV 481

Qy    339 AIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDT--------VNL 390
Db    482 GVAPGVG--VAPGIG------PGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGV 533

Qy    391 NG----------------------------THFNPLKKQGDEVKAGEL-------LCEFD 415
Db    534 PGLGVGAGVPGLGVGAGVPGFGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGA 593

Qy    416 IDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
Db    594 LAAAKAAKYGAAVPGVLGGLGALGGVGIPG-GVVGAGPAAAAAAAKAAAKA 643


RESULT   7
US-08-464-700-2
; Sequence 2, Application US/08464700
; Patent No. 6232458
;  GENERAL INFORMATION:
;    APPLICANT:  WEISS, ANTHONY S
;    APPLICANT:  MARTIN, STEPHEN L
;    TITLE OF INVENTION:  SYNTHETIC POLYNUCLEOTIDES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,700
;      FILING DATE:  7-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PL6520
;      FILING DATE:  22-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PL9661
;      FILING DATE:  28-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU93/00655
;      FILING DATE:  16-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  GHC3USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  733 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-464-700-2


  Query Match             4.6%;  Score 108.5;  DB 4;  Length 733;
  Best Local Similarity   20.6%;  Pred. No. 0.032;
  Matches  122;  Conservative   47;  Mismatches  227;  Indels  195;  Gaps   22;

Qy     16 ATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVL 75
Db    110 AGVKPGKVPGVGLPGV------YPGGVLP---------------GARFPGVG-------- 140

Qy     76 TLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL-VYSPIVITGLHQSF 134
Db    141 --VLPGVPTGAGVKPKAPGVGGAFA-GIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPY 197

Qy    135 PPIELELFNQGGGSFIFATASMANIAQG-----------AACLAVFFLAKSEKLKGLAGA 182
Db    198 TTGKLP-YGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAAAKAAAKFGAGAAGVLPGVGGA 256

Qy    183 SGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFL----- 237
Db    257 -GVPGVPGAI-PGIGGI---------AGVGTPAAAAAAAAA--AKAAKYGAAAGLVPGGP 303

Qy    238 -------------------------------------GVVSIDAPDMVMFL-------- 251
Db    304 GFGPGVVGVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARP 363

Qy    252 ---VCAVVTFFIAFGAAIAYGLYLVRRNG-----------------SIDPDATAAP-- 287
Db    364 GVGVGGIPTYGVGAGGFPGFGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAA 423

Qy    288 --------VPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASG-KLGSGV 338
Db    424 KAAKYGVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGV 483

Qy    339 AIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDT--------VNL 390
Db    484 GVAPGVG--VAPGIG-----PGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGV 535

Qy    391 NG----------------------------THFNPLKKQGDEVKAGEL-------LCEFD 415
Db    536 PGLGVGAGVPGLGVGAGVPGFGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGA 595

Qy    416 IDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
Db    596 LAAAKAAKYGAAVPGVLGGLGALGGVGIPG-GVVGAGPAAAAAAAKAAAKA 645


RESULT   8
US-08-836-325-2
; Sequence 2, Application US/08836325
; Patent No. 6110672
;  GENERAL INFORMATION:
;    APPLICANT:  Mandel, Gail
;    APPLICANT:  Halegoua, Simon
;    APPLICANT:  Borden, Laurence A.
;    TITLE OF INVENTION:  Peripheral Nervous System Specific
;    TITLE OF INVENTION:  Sodium Channels, DNA Encoding Therefor, Crystallization,
;    TITLE OF INVENTION:  X-ray Diffraction, Computer Molecular Modeling, Rational
;    TITLE OF INVENTION:  Drug Design, Drug Screening, and Methods of Making and Usin
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
;      STREET:  1100 New York Ave., N. W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,325
;      FILING DATE:  2-MAY-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  PCT/US95/14251
;      FILING DATE:  02-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/482,401
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/334,029
;      FILING DATE:  02-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0917.0240002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1011 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-836-325-2


  Query Match             4.6%;  Score 108;  DB 3;  Length 1011;
  Best Local Similarity   20.7%;  Pred. No. 0.06;
  Matches   91;  Conservative   67;  Mismatches  168;  Indels  114;  Gaps   20;

Qy      28 LFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL------ITPVLTLLLTG 81
           :| |   || :  |:: :: :: :   :||      :|  :::      |  |  :| ||
Db     553 IFDLVTNQA-FDITIMVLICLNMVTMMVEK------EGQTEYMDYVLHWINMVFIILFTG 605

Qy      82 FLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELEL 141
                |          | |     | |   ||   : ::|  :||  :   |    :|
Db     606 ECVLKLIS---------LRH-----YYF--TVG---WNILYFVVVILSIVGMFLAEMIEK 646

Qy     142 FNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNL 201
           :      | : ::  : : |    :  |     :||  |   :   | :: ||:| : |
Db     647 Y-------FVSPTLFRVIRLARIGRILRL-----IKGAKGIRTLLFALMMSLPALFNIGL 694

Qy     202 RLRWPFFI----GIGTAA---------------IGGALIALFNIKAVALGAAGFLGVVS 241
             |    ||    |:   |               | ::| || |   | |  | |  :
Db     695 LLFLVMFIYAIFGMSNFAYVKKEAGINDMFNFETFGNSMICLFQITTSA-GWDGLLAPIL 753

Qy     242 IDAP----------------------DMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGS 278
             ||                        :  |:   :::| :     ||    ::  | |
Db     754 NSAPPDCDPKKVHPGSSVEGDCGNPSVGIFYFVSYIIISFLVVVNMYIA----VILENFS 809

Qy     279 IDPDATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGV 338
           :  : :  |:     :   |  :|  |:|       : |    :||  ||:     :
Db     810 VATEESTEPLSEDDFEMFYEVWEKFDPDAT--------QFIEFCKLSD--FAAALDPPLL 859

Qy     339 AIVPTKGQLVS----PVSGKIVVAFPSGHAFAVRTKAEDGSNVDIL---MHIGFDTVNLN 391
             | | ||::     |||  :       ||  |   | |  :| |   |   | : | :
Db     860 IAKPNKVQLIAMDLPMVSGDRIHCLDILFAFTKRVLGE-GGEMDSLRSQMEERFMSANPS 918

Qy     392 GTHFNP----LKKQGDEVKA 407
             : |    ||:: :|| |
Db     919 KVSYEPITTTLKRKQEEVSA 938


RESULT   9
US-08-836-325-12
; Sequence 12, Application US/08836325
; Patent No. 6110672
;  GENERAL INFORMATION:
;    APPLICANT:  Mandel, Gail
;    APPLICANT:  Halegoua, Simon
;    APPLICANT:  Borden, Laurence A.
;    TITLE OF INVENTION:  Peripheral Nervous System Specific
;    TITLE OF INVENTION:  Sodium Channels, DNA Encoding Therefor, Crystallization,
;    TITLE OF INVENTION:  X-ray Diffraction, Computer Molecular Modeling, Rational
;    TITLE OF INVENTION:  Drug Design, Drug Screening, and Methods of Making and Using
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
;      STREET:  1100 New York Ave., N. W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,325
;      FILING DATE:  2-MAY-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/14251
;      FILING DATE:  02-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/482,401
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/334,029
;      FILING DATE:  02-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0917.0240002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1989 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-08-836-325-12


  Query Match             4.6%;  Score 108;  DB 3;  Length 1989;
  Best Local Similarity   20.7%;  Pred. No. 0.17;
  Matches   91;  Conservative   67;  Mismatches  168;  Indels  114;  Gaps   20;

Qy      28 LFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL------ITPVLTLLLTG 81
           :| |   || :  |:: :: :: :   :||      :|  :::      |  |  :| ||
Db    1508 IFDLVTNQA-FDITIMVLICLNMVTMMVEK------EGQTEYMDYVLHWINMVFIILFTG 1560

Qy      82 FLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELEL 141
                |          | |     | |   ||   : ::|  :||  :   |    :|
Db    1561 ECVLKLIS---------LRH-----YYF--TVG---WNILYFVVVILSIVGMFLAEMIEK 1601

Qy     142 FNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNL 201
           :      | : ::  : : |    :  |     :||  |   :   | :: ||:| : |
Db    1602 Y-------FVSPTLFRVIRLARIGRILRL-----IKGAKGIRTLLFALMMSLPALFNIGL 1649

Qy     202 RLRWPFFI----GIGTAA---------------IGGALIALFNIKAVALGAAGFLGVVS 241
             |    ||    |:   |               | ::| || |   | |  | |  :
Db    1650 LLFLVMFIYAIFGMSNFAYVKKEAGINDMFNFETFGNSMICLFQITTSA-GWDGLLAPIL 1708

Qy     242 IDAP----------------------DMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGS 278
             ||                        :  |:   :::| :     ||    ::  | |
Db    1709 NSAPPDCDPKKVHPGSSVEGDCGNPSVGIFYFVSYIIISFLVVVNMYIA----VILENFS 1764

Qy     279 IDPDATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGV 338
           :  : :  |:     :   |  :|  |:|       : |    :||  ||:     :
Db    1765 VATEESTEPLSEDDFEMFYEVWEKFDPDAT--------QFIEFCKLSD--FAAALDPPLL 1814

Qy     339 AIVPTKGQLVS----PVSGKIVVAFPSGHAFAVRTKAEDGSNVDIL---MHIGFDTVNLN 391
```

```
Db   1815 IAKPNKVQLIAMDLPMVSGDRIHCLDILFAFTKRVLGE-GGEMDSLRSQMEERFMSANPS 1873

Qy    392 GTHFNP----LKKQGDEVKA 407
Db   1874 KVSYEPITTTLKRKQEEVSA 1893


RESULT  10
US-08-836-325-10
; Sequence 10, Application US/08836325
; Patent No. 6110672
;  GENERAL INFORMATION:
;    APPLICANT:  Mandel, Gail
;    APPLICANT:  Halegoua, Simon
;    APPLICANT:  Borden, Laurence A.
;    TITLE OF INVENTION:  Peripheral Nervous System Specific
;    TITLE OF INVENTION:  Sodium Channels, DNA Encoding Therefor, Crystallization,
;    TITLE OF INVENTION:  X-ray Diffraction, Computer Molecular Modeling, Rational
;    TITLE OF INVENTION:  Drug Design, Drug Screening, and Methods of Making and Using
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
;      STREET:  1100 New York Ave., N. W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,325
;      FILING DATE:  2-MAY-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/14251
;      FILING DATE:  02-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/482,401
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/334,029
;      FILING DATE:  02-NOV-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0917.0240002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1984 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-836-325-10


  Query Match             4.5%;  Score 107;  DB 3;  Length 1984;
  Best Local Similarity   20.7%;  Pred. No. 0.22;
  Matches   91;  Conservative   66;  Mismatches  169;  Indels  114;  Gaps   20;

Qy      28 LFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFL------ITPVLTLLLTG 81
Db    1506 IFDLVTNQA-FDITIMVLICLNMVTMMVEK------EGQTEYMDYVLHWINMVFIILFTG 1558

Qy      82 FLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELEL 141
Db    1559 ECVLKLIS---------LRH-----YYF--TVGWNIFDFVVVILSIVGMFLA------EM 1596

Qy     142 FNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNL 201
Db    1597 IEK----YFVSPTLFRVIRLARIGRILRL-----IKGAKGIRTLLFALMMSLPALFNIGL 1647

Qy     202 RLRWPFFI----GIGTAA---------------IGGALIALFNIKAVALGAAGFLGVVS 241
Db    1648 LLFLVMFIYAIFGMSNFAYVKKEAGINDMFNFETFGNSMICLFQITTSA-GWDGLLAPIL 1706

Qy     242 IDAP----------------------DMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGS 278
Db    1707 NSAPPDCDPKKVHPGSSVEGDCGNPSVGIFYFVSYIIISFLVVVNMYIA----VILENFS 1762

Qy     279 IDPDATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGV 338
Db    1763 VATEESTEPLSEDDFEMFYEVWEKFDPDAT--------QFIEFCKLSD--FAAALDPPLL 1812

Qy     339 AIVPTKGQLVS----PVSGKIVVAFPSGHAFAVRTKAEDGSNVDIL---MHIGFDTVNLN 391
Db    1813 IAKPNKVQLIAMDLPMVSGDRIHCLDILFAFTKRVLGE-GGEMDSLRSQMEERFMSANPS 1871

Qy     392 GTHFNP----LKKQGDEVKA 407
Db    1872 KVSYEPITTTLKRKQEEVSA 1891


RESULT  11
5164180-6
;Patent No. 5164180
;    APPLICANT: Payne, Jewel;Sick, August J.
;    TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE
;AGAINST LEPIDOPTERAN PESTS
;    NUMBER OF SEQUENCES: 6
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/451,389
;      FILING DATE: 14-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 353,860
;      FILING DATE: 18-MAY-1989
;SEQ ID NO:6:
;      LENGTH: 1129
5164180-6


  Query Match             4.5%;  Score 106.5;  DB 6;  Length 1129;
  Best Local Similarity   22.4%;  Pred. No. 0.1;
  Matches   56;  Conservative   42;  Mismatches   97;  Indels   55;  Gaps   13;

Qy     249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
Db     380 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 437

Qy     295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
Db     438 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 496

Qy     337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
Db     497 DILRRTSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551

Qy     397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
Db     552 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 596

Qy     457 NVAKKEAVPA 466
Db     597 YIDRIEFVPA 606


RESULT  12
5188960-2
;Patent No. 5188960
```

```
;    APPLICANT: PAYNE, JEWEL;SICK, AUGUST J.
;    TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATE ACTIVE
;AGAINST LEPIDOPTERAN PESTS, AND GENES ENCODING NOVEL
;LEPIDOPTERAN-ACTIVE TOXINS
;    NUMBER OF SEQUENCES: 8
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/451,261
;      FILING DATE: 14-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 371,955
;      FILING DATE: 27-JUN-1989
;SEQ ID NO:2:
;      LENGTH: 1179
5188960-2


  Query Match             4.5%;  Score 106.5;  DB 6;  Length 1179;
  Best Local Similarity   22.4%;  Pred. No. 0.11;
  Matches   56;  Conservative   42;  Mismatches   97;  Indels   55;  Gaps   13;

Qy    249 MFLVCAVVTFFIAFGAAIAYGLYLVRRNGSID------PDATAAPVPAG--------TTK 294
          :|::      | :  | :   :|  |: |::|      |   : |  ||         |
Db    380 LFVLDGTEFSFASLTADLPSTIY--RQRGTVDSLDVIPPQDNSVPARAGFSHRLSHVTML 437

Qy    295 AEA-------EAP--------AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
          ::|        ||         |||||    | | | ||| ::| | | :  :    |  |
Db    438 SQAAGAVYTLRAPTFSWRHRSAEFSNLIPSSQITQIPLT-KSINLGSGTSVVKGPGFTGG 496

Qy    337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFN 396
           :    : ||: :    :: :  |    : || :    |  ::  |   |   :|  :|:
Db    497 DILRRTSPGQISTL---RVTITAPLSQRYRVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551

Qy    397 PLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLL 456
              |  :::|         : : ||:  |||    |:     : |        :| | :
Db    552 ATMSSGGNLQSG---------SFRTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 596

Qy    457 NVAKKEAVPA 466
           : : | |||
Db    597 YIDRIEFVPA 606


RESULT  13
US-08-804-227C-3
; Sequence 3, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1864 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-3


  Query Match             4.5%;  Score 106.5;  DB 2;  Length 1864;
  Best Local Similarity   21.7%;  Pred. No. 0.22;
  Matches   94;  Conservative   45;  Mismatches  159;  Indels  135;  Gaps   20;

Qy      7 SLVNGYDVAATMAAGEMPMW-----SLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHK 61
          :|:  |  | |    | |:|     | |   | ||  | :  || :          |
Db    346 ALLATYGTART---AERPLWLGSLKSNIGHTQAAAGVAGVIKMVLAMR----------HG 392

Qy     62 RL----------------KGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQG 105
          ||                ||    |  ||                   | |:       |
Db    393 RLPRTLHVDRPTTRVDWEKGGVRLLTEPV-----------------PWPGEAGEPRRAG 434

Qy    106 LYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACL 165
          :  ||            : :|:  :    ||      : ||::   | |     :| | |
Db    435 VSSFGA-------SGTNAHVVLESVPAGEPPAAGRPEDTGGAW---TVS----GRGPAAL 480

Qy    166 AVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALI---A 222
              : | |    :|  | |                  | ||| : |||
Db    481 RAQAARLYDALTGTGTGTGQGAGQG-----------------AGPGTAEVAGALAHART 522

Qy    223 LFNIKAVALGA------AGFLGVVSIDAPDMVMFLVCAVV----TFFIAFG-----AAIA 267
           |  :|| ||        ||   :  : |   :    |       | |:  |     |
Db    523 AFRHRAVVLGGNRAELLAGLRELAEEEHPGPRVVTGTAPATERRTAFLFSGQGSQRAGSG 582

Qy    268 YGLYLVRRN----------GSIDPD----------ATAAPVPAGTTKAEAEAPAEFSN 305
           |||  ||:            :::|              :|      : |   || |:
Db    583 RGLY--RRHPVFARALDEVCAALEPHLHRPLRDLMFAEPGSPEAEPLDRTEFTQPALFAL 640

Qy    306 DSTIIQAP----LTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFPS 361
          : : :      |  ||:   || :   |     :||  :|   :||:   |::: | |:
Db    641 QTALFRLAEHHGLRAEALCGHSVGEIAAAH---AAGVLTLPDAARLVA-ARGRLMQALPA 696

Qy    362 GHAF-AVRTKAED 373
          | |  |:|  ||:
Db    697 GGAMAALRATAEE 709


RESULT  14
US-08-801-344-9
; Sequence 9, Application US/08801344
; Patent No. 6087140
;  GENERAL INFORMATION:
;    APPLICANT:  Cameron, Douglas C.
;    APPLICANT:  Shaw, Anita J.
;    APPLICANT:  Altaras, Nedim E.
;    TITLE OF INVENTION:  MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL
;    TITLE OF INVENTION:  FROM SUGAR
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DeWitt Ross & Stevens S.C.
;      STREET:  8000 Excelsior Drive, Suite 401
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53717-1914
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,344
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REFERENCE/DOCKET NUMBER:  09820.037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-831-2100
;      TELEFAX:  608-831-2106
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Pyridine nucleotide transhydrogenase, subunit
;      ORGANISM:  B
US-08-801-344-9


  Query Match             4.5%;  Score 105.5;  DB 3;  Length 462;
  Best Local Similarity   21.1%;  Pred. No. 0.032;
  Matches   91;  Conservative   66;  Mismatches  123;  Indels  151;  Gaps   24;

Qy      27 SLFGLDVAQAGY------------------QGTVLPVLV--------VSWILATIEKFL 59
           ::|| |    |:                   : | :| ||        :: :|     :|
Db      48 TIFGPDTGNVGWILLAMVIGGAIGIRLAKKVEMTEMPELVAILHSFVGLAAVLVGFNSYL 107

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
           |    | |  |:   || :  |    | || |: : | |:|             | | |
Db     108 HHD-AGMAPILVNIHLTEVFLG----IFIG-AVTFTGSVVAF------------GKLCG 148

Qy     120 LVYS-PIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKG 178
           : | |:::   |:                    ::| :      | ||    |  |:
Db     149 KISSKPLMLPNRHK-------------------MNLAALVVSFLLLIVFVRTDSVGLQ- 187

Qy     179 LAGASGVSAVLGITEPA-IFGVNLRLRWPFFIGIGTAAIGGA----LIALFN-IKAVALG 232
                 | |:| :|  | :|| :|            |:|||    :::: |     |
Db     188 ------VLALLIMTAIALVFGWHL-----------VASIGGADMPVVVSMLNSYSGWAAA 230

Qy     233 AAGFL---------GVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDA 283
           ||||:         | :   : :: :::|  :    :| : || |          |  :
Db     231 AAGFMLSNDLLIVTGALVGSSGAILSYIMCKAMN--RSFISVIAGGF-------GTDGSS 281

Qy     284 TAAPVPAGT-TKAEAEAPAE-FSNDSTIIQAPLTGEAIA-----LSSVSDAMFASG---K 333
           |      |   :  || ||   |  ::|  |  | |:|     :: ::: : | |   :
Db     282 TGDDQEVGEHREITAEETAELLKNSHSVIITPGYGMAVAQAQYPVAEITEKLRARGINVR 341

Qy     334 LGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKA-------------EDGSNVDI 379
           |            : ||:|::      ||   :  :|              :| :: |
Db     342 FG-----------IHPVAGRL-----PGHMNVLLAEAKVPYDIVLEMDEINDDFADTDT 384

Qy     380 LMHIGF-DTVN 389
           :: ||  ||||
Db     385 VLVIGANDTVN 395


RESULT  15
US-09-498-599-9
; Sequence 9, Application US/09498599
; Patent No. 6303352
;  GENERAL INFORMATION:
;    APPLICANT:  Cameron, Douglas C.
;    APPLICANT:  Shaw, Anita J.
;    APPLICANT:  Altaras, Nedim E.
;    TITLE OF INVENTION:  MICROBIAL PRODUCTION OF
;    TITLE OF INVENTION:  1,2-PROPANEDIOL FROM SUGAR
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DeWitt Ross & Stevens S.C.
;      STREET:  8000 Excelsior Drive, Suite 401
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53717-1914
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/498,599
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REFERENCE/DOCKET NUMBER:  09820.037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-831-2100
;      TELEFAX:  608-831-2106
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  462 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Pyridine nucleotide transhydrogenase, subunit
;      ORGANISM:  B
US-09-498-599-9


  Query Match             4.5%;  Score 105.5;  DB 4;  Length 462;
  Best Local Similarity   21.1%;  Pred. No. 0.032;
  Matches   91;  Conservative   66;  Mismatches  123;  Indels  151;  Gaps   24;

Qy      27 SLFGLDVAQAGY------------------QGTVLPVLV--------VSWILATIEKFL 59
           ::|| |    |:                   : | :| ||        :: :|     :|
Db      48 TIFGPDTGNVGWILLAMVIGGAIGIRLAKKVEMTEMPELVAILHSFVGLAAVLVGFNSYL 107

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
           |    | |  |:   || :  |    | || |: : | |:|             | | |
Db     108 HHD-AGMAPILVNIHLTEVFLG----IFIG-AVTFTGSVVAF------------GKLCG 148

Qy     120 LVYS-PIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKG 178
           : | |:::   |:                    ::| :      | ||    |  |:
Db     149 KISSKPLMLPNRHK-------------------MNLAALVVSFLLLIVFVRTDSVGLQ- 187

Qy     179 LAGASGVSAVLGITEPA-IFGVNLRLRWPFFIGIGTAAIGGA----LIALFN-IKAVALG 232
                 | |:| :|  | :|| :|            |:|||    :::: |     |
Db     188 ------VLALLIMTAIALVFGWHL-----------VASIGGADMPVVVSMLNSYSGWAAA 230

Qy     233 AAGFL---------GVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDA 283
           ||||:         | :   : :: :::|  :    :| : || |          |  :
Db     231 AAGFMLSNDLLIVTGALVGSSGAILSYIMCKAMN--RSFISVIAGGF-------GTDGSS 281

Qy     284 TAAPVPAGT-TKAEAEAPAE-FSNDSTIIQAPLTGEAIA-----LSSVSDAMFASG---K 333
           |      |   :  || ||   |  ::|  |  | |:|     :: ::: : | |   :
Db     282 TGDDQEVGEHREITAEETAELLKNSHSVIITPGYGMAVAQAQYPVAEITEKLRARGINVR 341

Qy     334 LGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKA-------------EDGSNVDI 379
           |            : ||:|::      ||   :  :|              :| :: |
Db     342 FG-----------IHPVAGRL-----PGHMNVLLAEAKVPYDIVLEMDEINDDFADTDT 384

Qy     380 LMHIGF-DTVN 389
           :: ||  ||||
Db     385 VLVIGANDTVN 395
```

Search completed: September 25, 2002, 06:00:58
Job time: 2334 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2002, 05:55:04 ; Search time 45.29 Seconds
(without alignments)
992.930 Million cell updates/sec

Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MAMVFPSLVNGYDVAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1058.5	44.8	651	2	S44257	phosphotransferase
2	953.5	40.4	664	1	B32243	phosphotransferase
3	759.5	32.1	627	2	H96951	fusion, PTS system
4	701	29.7	480	2	S39978	scrA protein - Sta
5	695.5	29.4	630	2	H83686	PTS system, beta-g
6	695	29.4	617	2	AB1167	phosphotransferase
7	674	28.5	633	2	AC1436	PTS system, beta-g
8	673.5	28.5	636	2	C83724	PTS system, beta-g
9	673	28.5	632	2	S68599	phosphotransferase
10	673	28.5	655	2	C95220	trehalose PTS syst
11	667	28.2	634	2	AD1078	PTS system, beta-g
12	664	28.1	627	2	F95200	PTS system IIABC c
13	660	27.9	627	2	E98067	phosphotransferase
14	659	27.9	479	2	F82432	PTS system, sucros
15	659	27.9	480	2	D90038	PTS system, sucros
16	658	27.8	705	2	A99084	phosphotransferase
17	657	27.8	609	2	I40406	beta-glucoside per
18	646	27.3	479	2	JQ0781	sucrose uptake pro
19	645	27.3	609	2	T47097	hypothetical prote
20	640	27.1	631	2	B42603	beta-glucoside-spe
21	628	26.6	617	2	AC1421	beta-glucoside-spe
22	612	25.9	628	2	D97073	PTS system, beta-g
23	588.5	24.9	636	2	D86807	hypothetical prote
24	580	24.5	618	2	AC1204	phosphotransferase
25	578.5	24.5	612	2	A97935	hypothetical prote
26	570.5	24.1	612	2	B95067	hypothetical prote
27	570	24.1	470	2	C69725	phosphotransferase
28	569.5	24.1	470	2	H83926	PTS system, trehal
29	567.5	24.0	625	2	C25977	phosphotransferase
30	509	21.5	640	2	AB1423	beta-glucoside-spe
31	496	21.0	475	2	C89813	hypothetical prote
32	489	20.7	456	2	S62331	phosphotransferase
33	487	20.6	458	2	H83881	PTS system, sucros
34	473	20.0	455	1	WQEBST	phosphotransferase
35	467	19.8	372	2	I39868	sac operon regulat
36	408	17.3	494	2	AG1231	PTS system trehalo
37	406	17.2	459	2	JU0293	levansucrase synth
38	396.5	16.8	494	2	AF1585	PTS system trehalo
39	390	16.5	460	2	A39938	phosphotransferase
40	367	15.5	473	2	C65236	phosphotransferase
41	367	15.5	473	2	A98281	trehalose specific
42	367	15.5	483	2	AI0449	protein-Npi-phosph
43	360	15.2	473	2	A86122	PTS system enzyme
44	348.5	14.7	681	2	A89781	hypothetical prote
45	345	14.6	665	2	B96970	PTS enzyme II, ABC

ALIGNMENTS

RESULT 1
S44257
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pediococcus pen
C;Species: Pediococcus pentosaceus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999
C;Accession: S44257
R;Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.
submitted to the EMBL Data Library, April 1994
A;Description: The sucrose and raffinose operons of Pediococcus pentosaceus PPE1.0.
A;Reference number: S44252
A;Accession: S44257
A;Molecule type: DNA
A;Residues: 1-651 <LEE>
A;Cross-references: EMBL:Z32771; NID:g493728; PIDN:CAA83668.1; PID:g475968
C;Genetics:
A;Gene: scrA
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera
C;Keywords: phosphotransferase
F;488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 44.8%; Score 1058.5; DB 2; Length 651;
Best Local Similarity 46.7%; Pred. No. 1.2e-65;
Matches 221; Conservative 78; Mismatches 151; Indels 23; Gaps 5;

```
Qy      1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          | || ||||||| || |||||:|  |::||| |||||||| ||||| |::||||:||| |
Db    194 MIMVLPSLVNGYSVATTMAAGKMVYWNVFGLHVAQAGYQGQVLPVLGVAFILATLEKFFH 253

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          | :||  ||  ||:  :::|||||||  :|| :| | | | :|| |||:  | :|  :|||
Db    254 KHIKGAFDFTFTPMFAIVITGFLTFTIVGPVLRTVSDALTNGLVGLYNSTGWIGMGIFGL 313

Qy    121 VYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
          :|| ||||||||:|| || :|       ||||||  |||||| |||| ||:||  ||:|
Db    314 LYSAIVITGLHQTFPAIETQLLANVAKTGGSFIFPVASMANIGQGAATLAIFFATKSQKQ 373

Qy    177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
          | |  ::||||:||||||||||||:::::||     : |  | : ||:: :||:| |
Db    374 KALTSSAGVSALLGITEPAIFGVNLKMKFPFVFAAIASGIASAFLGLFHVLSVAMGPASV 433

Qy    237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAE 296
          :| :|| :  :  |:: ||::| :||     |     :|    | |   :| | |  |
Db    434 IGFISIASKSIPAFMLSAVISFVVAFIPTFIY----AKRTLGDDRDQVKSPAPTSTV--- 486

Qy    297 AEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIV 356
              : :  || ||:|| : :|  |:| :|::  :| | ||||:  |:|:|  | |
Db    487 ------INVNDEIISAPVTGASESLKQVNDQVFSAEIMGKGAAIVPSSDQVVAPADGVIT 540

Qy    357 VAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDI 416
          | : | ||: ::| |      :||:|:| ||||||| ||    ::|| |  |:||  |||
Db    541 VTYDSHHAYGIKTTA----GAEILIHLGLDTVNLNGEHFTTNVQKGDTVHQGDLLGTFDI 596
```

```
Qy    417 DAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANL--LNVAKKEAVPAT 467
          |:||| |: |  ::|:|      |    :  ::||  |  |      :| ||
Db    597 AALKAANYDPTVMLIVTNTANYANVERLKVTNVQAGEQLVALTAPAASSVAAT 649


RESULT   2
B32243
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus muta
C;Species: Streptococcus mutans
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: B32243
R;Sato, Y.; Poy, F.; Jacobson, G.R.; Kuramitsu, H.K.
J. Bacteriol. 171, 263-271, 1989
A;Title: Characterization and sequence analysis of the scrA gene encoding enzyme II(scr)
A;Reference number: A32243; MUID:89123027
A;Accession: B32243
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <SAT>
A;Cross-references: GB:M22711; NID:g153799; PIDN:AAA26971.1; PID:g153801
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase
C;Keywords: phosphoprotein; phosphotransferase
F;511-664/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol


  Query Match             40.4%;  Score 953.5;  DB 1;  Length 664;
  Best Local Similarity   42.6%;  Pred. No. 2.3e-58;
  Matches  205;  Conservative   89;  Mismatches  156;  Indels   31;  Gaps    9;

Qy      1 MAMVFPSLVNGYDVAATMAAGEMP-------MWSLFGLDVAQAGYQGTVLPVLVVSWILA 53
          | || |   |   :||     :             |::||| | || |   |:|||  |:|:
Db    194 MIMVAPGAANIIGLAANAPISKAATIGAYTGFWNIFGLHVTQASYTYQVIPVLVAVWLLS 253

Qy     54 TIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPV 113
           :||| ||||    ||  ||:|::::|||||| ||| |: | | | :|:  |||  | :
Db    254 ILEKFFHKRLPSAVDFTFTPLLSVIITGFLTFIVIGPVMKEVSDWLTNGIVWLYDTTGFL 313

Qy    114 GGLLFGLVYSPIVITGLHQSFPPIELEL---FNQG---GSFIFATASMANIAQGAACLAV 167
          |  :|| :|||:|:|||||||| || :|    |  |    | ||| ||||||:|||||  |:
Db    314 GMGVFGALYSPVVMTGLHQSFPAIETQLISAFQNGTGHGDFIFVTASMANVAQGAATFAI 373

Qy    168 FFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIK 227
          :|| | :|:|||: :|||||:|||||||:|||||: |:|||  :  :|   |:  |  :
Db    374 YFLTKDKKMKGLSSSSGVSALLGITEPALFGVNLKYRFPFFCALIGSASAAAIAGLLQVV 433

Qy    228 AVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAP 287
          ||:||:|||||  :|| |  :  ::|| :::| |||      ||      :  ::|  |  |
Db    434 AVSLGSAGFLGFLSIKASSIPFYVVCELISFAIAFAVTYGYG-----KTKAVDVFAAEAA 488

Qy    288 VPAGTTKAEAEAPAEFSNDST-------IIQAPLTGEAIALSSVSDAMFASGKLGSGVAI 340
          |    : : | | | :: :         :: ||| |||: |:||:| :|:|  :| |:||
Db    489 VEEAIEEVQ-EIPEEAASAANKAQVTDEVLAAPLAGEAVELTSVNDPVFSSEAMGKGIAI 547

Qy    341 VPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKK 400
          |:   : :|| | : :|| :|||: :  |:::|:  :||:||| |||:: |  |     :
Db    548 KPSGNTVYAPVDGTVQIAFDTGHAYGI--KSDNGA--EILIHIGIDTVSMEGKGFEQKVQ 603

Qy    401 QGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTY-GLGEIEAGANLLNVA 459
             ::| |::|  || | |  || : ||  :|:|      | |    | :  | :|| |
Db    604 ADQKIKKGDVLGTFDSDKIAEAGLDNTTMFIVTNTADYASVETLASSGTVAVGDSLLEVK 663

Qy    460 K 460
          |
Db    664 K 664


RESULT   3
H96951
fusion, PTS system, beta-glucosides specific IIABC component [imported] - Clostridium ac
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: H96951
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: H96951
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-627 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78403.1; PID:g15023277; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0423
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera


  Query Match             32.1%;  Score 759.5;  DB 2;  Length 627;
  Best Local Similarity   36.7%;  Pred. No. 5.6e-45;
  Matches  163;  Conservative   80;  Mismatches  176;  Indels   25;  Gaps    5;

Qy      3 MVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKR 62
          |: | | | :    |:  |     ::|||::  ||||||||:|:  |::: ||| | |
Db    186 MIHPDLQNAW----TLGEGIKHTINIFGLNIGMVGYQGTVLPILISVWVMSYIEKGLRKI 241

Qy     63 LKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVY 122
          :    | |:|| |||::|||   : |||  |:||| :: ||| ||:  |   |:||| :|
Db    242 VPEALDILLTPFLTLMITGFFAMVVIGPGGRFVGDEISLGLQTLYNTTGFFSGVLFGGLY 301

Qy    123 SPIVITGLHQSFPPIELELFNQGG---SFIFATASMANIAQGAACLAVFFLAKSEKLKGL 179
          | |||||:| ||  ||  |         :|:    ||||:||| | |||:|  : :|:| :
Db    302 SLIVITGIHHSFHAIEAGLLANPAIHKNFLLPIWSMANVAQGGAALAVYFKTRDKKMKSI 361

Qy    180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGV 239
          |  :  | :|||||||||||||||   ||  |    ||||  |    :   |:|  |  |:
Db    362 AAPASFSCLLGITEPAIFGVNLRYTKPFIAGALGGAIGGGYIVFTKVAMTAVGVTGIPGI 421

Qy    240 VSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEA 299
            :     : :::  :    :|||  |    :  |  :    : |      |        | |:
Db    422 AIVKQGSFLNYIIAMI----LAFGGAFIIAMVLGIKEEITEEDLNKETVNKDIKVEEVES 477

Qy    300 PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 359
                  : :|: |: : | :| |  || | :|  |: : |  |::|||: | :|  |
Db    478 ----------VVSPVNGKVVLLKNVPDKTFAEGLIGDGIGVDPEDGEVVSPIDGTVVHVF 527

Qy    360 PSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAI 419
          : || |:::|    : |::|:||| ||| : |  |      |:||| |: | :||:| :
Db    528 ETKHAIAMKSK----NGVEMLIHIGIDTVKMEGNGFKSFINDGEEVKKGDKLIQFDLDLV 583

Qy    420 KAAGYEVTTPIVVSNYKKTGPVNT 443
          |          :|:|::  | ||:
Db    584 KEKAVSPIVLTIVTNHEDMGFVNS 607


RESULT   4
S39978
scrA protein - Staphylococcus xylosus
C;Species: Staphylococcus xylosus
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
C;Accession: S39978
R;Wagner, E.; Goetz, F.; Brueckner, R.
Mol. Gen. Genet. 241, 33-41, 1993
A;Title: Cloning and characterization of the scrA gene encoding the sucrose-specific
A;Reference number: S39976; MUID:94049686
A;Accession: S39978
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-480 <WAG>
A;Cross-references: EMBL:X69800; NID:g407905; PIDN:CAA49461.1; PID:g407908
C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II


  Query Match             29.7%;  Score 701;  DB 2;  Length 480;
  Best Local Similarity   51.4%;  Pred. No. 4.6e-41;
```

```
  Matches  146;  Conservative   49;  Mismatches   81;  Indels    8;  Gaps    4;

Qy       1 MAMVFPSLVNGYDVAATMAAG-EMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
           | :| | |:: ||    : || |:| |:||||:: | |||| |||:|| ::||||||| |
Db     194 MILVHPELMSAYDYPKALEAGKEIPHWNLFGLEINQVGYQGQVLPMLVATYILATIEKGL 253

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
            | :    | |:||:| :| |||:||  :||  | :|  |: ||  ||:||| :|||:||
Db     254 RKVIPTVLDNLLTPLLAILSTGFITFSFVGPLTRTLGYWLSDGLTWLYEFGGAIGGLIFG 313

Qy     120 LVYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAK-SE 174
           |:|:||||||:| ||  || :|     : |||||  |:|:||||||| || ||: | ::
Db     314 LLYAPIVITGMHHSFIAIETQLIADSSSTGGSFIFPIATMSNIAQGAAALAAFFIIKENK 373

Qy     175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
           ||||:| |:||||:|||||||:|||||:||:||   |  : || | || | :||:||| |
Db     374 KLKGVASAAGVSALLGITEPAMFGVNLKLRYPFIGAIVGSGIGSAYIAFFKVKAIALGTA 433

Qy     235 GFLGVVSIDAPD--MVMFLVCAVVTFFIAFGAAIAYGLYLVRRN 276
           |  | :||   :   : : :  :: | :|||   |       ||
Db     434 GIPGFISISGQNNGWLHYGIAMIIAFIVAFGVTYALSYRKKYRN 477


RESULT   5
H83686
PTS system, beta-glucoside-specific enzyme II, ABC component BH0296 [imported] - Bacillu
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: H83686
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83686
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-630 <STO>
A;Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB04015.1; GSPDB:GN0
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0296
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase


  Query Match             29.4%;  Score 695.5;  DB 2;  Length 630;
  Best Local Similarity   32.8%;  Pred. No. 1.5e-40;
  Matches  152;  Conservative  108;  Mismatches  182;  Indels   21;  Gaps    5;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHK 61
           |:::||:|  :| |  :        : ||: |    |  || |:|:  : :: :||| :|
Db     185 ALIYPSIVELHDSAIDV--------TFFGIPVVLMNYTSTVFPILLAVFAMSYVEKFCNK 236

Qy      62 RLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLV 121
           ::       :||:: |::   :| | :||   ::|: :|  :| :: |   : | :   :
Db     237 KIHEAVKNFVTPLILLVVIVPVTLIILGPIGVYLGNGIASVIQEIFTFSPVLAGAIVAGI 296

Qy     122 YSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           :  :|| |:|    || |  :|   | | |: |  :|  | | |    |::||| ||
Db     297 WQVLVIFGIHWGIIPIILNNLSVRGEDVIKAVAAPAVFSQAGAALGVMLRTKNKKLKALA 356

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
           |:: ::|: ||||||::|| | |: || : : :||:|||::   :   ||| || | | :
Db     357 GSTSITALFGITEPAVYGVTLPLKKPFIMAVISAAVGGAIVGHYGSVAVAPGAPGLLTIP 416

Qy     241 SI---DAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEA 297
                |    | |::  :::| :|     |   |:|     :| : | :       : :
Db     417 IFYPEDGRGFVAFVIAIIISFVLA-----AVLTYIVGFKDPVDDEDTLSNESGSENEVKR 471

Qy     298 EAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVV 357
           |   :  : |  |::|| || : |: | | :|:|| :| |||: | :|:||:|::| :
Db     472 EDDKKEPSASEEIKSPLKGEVVPLTEVQDHVFSSGAMGKGVAVRPKEGRLVAPINGTVTS 531

Qy     358 AFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDID 417
           | : ||  : :     :  :| :|:| ||| | | ||    :|||| ||::| |||::
Db     532 LFETKHAIGITS----DNGTEIFIHVGIDTVQLKGEHFTSFIEQGDEVAAGDVLLEFDVE 587

Qy     418 AIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           | ||||:| ||::::| |:   | |    |: :   |::| |
Db     588 RITAAGYDVITPVLITNAKQFSNVQTTDKREVTSEDLLIHVIK 630


RESULT   6
AB1167
phosphotransferase system (PTS) beta-glucoside-specific enzyme IIABC component homolo
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AB1167
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1167
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-617 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98816.1; PID:g16410127; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0738
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera


  Query Match             29.4%;  Score 695;  DB 2;  Length 617;
  Best Local Similarity   34.8%;  Pred. No. 1.6e-40;
  Matches  154;  Conservative   82;  Mismatches  162;  Indels   44;  Gaps    6;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHK 61
           |:|:|:::| ::  | :   ::|        |    | :|:|:::  | |: :|:||:
Db     185 ALVYPTMINLFNEGAHITFLQIP--------VVLMSYSFSVIPIILAVWFLSILERFLNS 236

Qy      62 RLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLV 121
           ::   |   :||:: |:|   |||:| ||   ::   || |   :|:    | |   |
Db     237 KIHEAAKTFLTPMICLMLIVPLTFLAFGPLGTFISQGLASGYTFIYNLSPIVAGAFMGAF 296

Qy     122 YSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           :  :|| |:|  | || :  |   |   : |    :| ||  | | ||   |  ::| :|
Db     297 WQVLVIFGIHWGFVPIMINNLSRYGRDTMIAMVGPSNFAQAGASLGVFLKTKKPEVKAIA 356

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFL--- 237
           |:: ::   |||||:|:|| |: : || |     |||||::       |    | |
Db     357 GSAALTGFFGITEPSIYGVTLKYKKPFVIASIAGAIGGAIVGAAGSSGAANAIPGILTLP 416

Qy     238 -----GVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGT 292
                | |       | ::: |: |:|  :   :| |:                   | |
Db     417 IFIGKGFVGFILGIAVAYILSAIGTYFFGYKDEMADGI-----------------APT 457

Qy     293 TKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVS 352
           ||   |   |    : :| :|: |  : |: | |  |::| || |||||| :|:|:|||:
Db     458 TKEAKETGVE----AEVIVSPIRGNIVPLNEVKDEAFSAGLLGKGVAIVPQEGKLISPVN 513

Qy     353 GKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLC 412
           | |  |||:||| :|:       |:||:|:|||| ||| :|  |  ||| |  |: |
Db     514 GTIETAFPTGHAIGIRS----DKGVEILLHVGFDTVQLNGKYFKLLVAQGDRVLVGQALL 569

Qy     413 EFDIDAIKAAGYEVTTPIVVSN 434
           |||:|||| ||::|||||:|
Db     570 EFDLEAIKADGYDITTPIVVTN 591


RESULT   7
AC1436
PTS system, beta-glucosides specific enzyme IIABC homolog lin0026 [imported] - Lister
```

```
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AC1436
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1436
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-633 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95259.1; PID:g16412447; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0026
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

  Query Match             28.5%;  Score 674;  DB 2;  Length 633;
  Best Local Similarity   36.2%;  Pred. No. 4.6e-39;
  Matches  161;  Conservative   75;  Mismatches  179;  Indels   30;  Gaps   10;

Qy       2 AMVFPSL--VNGYDVAATMAAG---EMPMWSLF-GLDVAQAGYQGTVLPVLVVSWILATI 55
           |:|:|:|  ::  |   |: ||   | |:   | |: |    |  :|:|::: ::  : :
Db     181 ALVYPTLAGISAGDPIYTLFAGTIFESPIHVTFLGIPVILMSYASSVIPIILATYFGSKV 240

Qy      56 EKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGG 115
           ||   | :      : |  |||:   :||: |||   | | :|  |   :|:    | |
Db     241 EKGFKKIIPDVIKTFVVPFCTLLVVVPITFLVIGPIATWAGQLLGAGTIWVYNLSPVVAG 300

Qy     116 LLFGLVYSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSE 174
           |: |  :   || |||    |: :  |   |   : |    |: ||  | |||||  ::: 
Db     301 LILGGFWQVFVIFGLHWGLIPVAINNLTTLGSDPVLAMMFGASFAQIGAVLAVFFKTRNK 360

Qy     175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
           |:| |:  : :| : |:|||||:|| | |: || :     |:|| :|   : |   :|
Db     361 KIKSLSIPAFISGIFGVTEPAIYGVTLPLKKPFIMSCIAGAVGGGIIGFVSAKVYIMGGL 420

Qy     235 GFLGVVSIDAP-----DMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVP 289
           |  |: :   |         ::| |:|  ||        |  |    |  ||   |  |
Db     421 GIFGLPNFFQPGAGITSAFWWVVIAIVISFI-------LGFILTYVAGFKDP---AEAVV 470

Qy     290 AGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVS 349
             |  | |   |       | ||: || : |: | |  |:|| || ||||:|| |::|: 
Db     471 EETNVTEGETLIE----RETIPAPVVGEIVTLADVKDEAFSSGALGKGVAIIPTVGRVVA 526

Qy     350 PVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGE 409
           | :|  :   ||:|||  : |   ||:  ::|:|||  ||| | |  |    |||| :: |: 
Db     527 PAAGTVTTIFPTGHAIGITT--NDGA--EVLIHIGMDTVQLEGKFFTAHVKQGDVIEKGQ 582

Qy     410 LLCEFDIDAIKAAGYEVTTPIVVSN 434
           || ||||: ||||||:||||:||:|
Db     583 LLTEFDIEGIKAAGYDVTTPVVVTN 607


RESULT   8
C83724
PTS system, beta-glucoside-specific enzyme II, ABC component bglP [imported] - Bacillus
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: C83724
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: C83724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-636 <STO>
A;Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04314.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: bglP
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

  Query Match             28.5%;  Score 673.5;  DB 2;  Length 636;
  Best Local Similarity   34.0%;  Pred. No. 5e-39;
  Matches  163;  Conservative   78;  Mismatches  193;  Indels   45;  Gaps    9;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLF-------------GLDVAQAGYQGTVLPVLV 47
           ::|:|:||        :  || |:::||               |: |    |  :|:|::: 
Db     183 SLVYPTLV-------VLTEGE-PLYTLFTGTIFESPVHITFLGIPVILMSYATSVIPIIL 234

Qy      48 VSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLY 107
           ::  : :|  | | :       : |  |||:   |||| |||   | | :|      :|
Db     235 AAYFASKVEARLRKIIPDVVKTFLVPFFTLLIVVPLTFIVIGPIATWAGQLLGQFTLWVY 294

Qy     108 DFGGPVGGLLFGLVYSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLA 166
           :    : |   |  :   || |||    || :  |  ||   : |    |: ||  |  |
Db     295 NLSPIIAGAFLGGFWQVFVIFGLHWGLIPIAINNLVVQGSDPVLAMVFAASFAQIGAVAA 354

Qy     167 VFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNI 226
           |:   | :|:|  |:  : :| : |:|||||:|| | |: || |    ||:|||:| ||
Db     355 VWLKIKQQKVKTLSVPAFISGIFGVTEPAIYGVTLPLKRPFIISCIAAAVGGAIIGLFRS 414

Qy     227 KAVALGAAGFLGVVSI----DAPDMVMF-LVCAVVTFFIAFGAAIAYGLYLVRRNGSIDP 281
           :   :|  |  |: |      |  |   : :| |||  |:  |  : |   |    |:  | :
Db     415 QGYIIGGLGIFGIPSFLHPADGMDAGFWGIVIAVVVAFV-LGFILTYLFGLKSGNASDEQ 473

Qy     282 DATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIV 341
             | |    || : |              | :|  |  | || : |  |:|| || |:||
Db     474 TETKAHTSTGTGEKEE------------ISSPFNGSVITLSEIKDEAFSSGALGEGIAIE 521

Qy     342 PTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQ 401
           |::|:| |||||  :   :|: ||  : |         ::|:||| ||| |:|  |      |
Db     522 PSEGKLFSPVSGMVTALYPTHHALGITT----DRGAELLIHIGLDTVQLDGKFFTAHTIQ 577

Qy     402 GDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           | :|: |:|| ||||  |||||:|:|:|:|: | :      :: ||  || : :
Db     578 GAQVEKGDLLIEFDIKEIKAAGYAVTTPVIVTNHKQYGQLFLTDKQQVNAGDRLLELTR 636


RESULT   9
S68599
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus s
N;Alternate names: sucrose-specific enzyme II
C;Species: Streptococcus sobrinus
A;Variety: strain 6715
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
C;Accession: S68599
R;Chen, Y.Y.M.; Lee, L.N.; LeBlanc, D.J.
Infect. Immun. 61, 2602-2610, 1993
A;Title: Sequence analysis of scrA and scrB from Streptococcus sobrinus 6715.
A;Reference number: S68598; MUID:93273516
A;Accession: S68599
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-632 <CHE>
A;Cross-references: EMBL:L06791
C;Genetics:
A;Gene: scrA
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera
C;Keywords: phosphotransferase; sugar transport system
F;480-632/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

  Query Match             28.5%;  Score 673;  DB 2;  Length 632;
  Best Local Similarity   36.4%;  Pred. No. 5.4e-39;
  Matches  168;  Conservative   76;  Mismatches  197;  Indels   20;  Gaps    8;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
```

```
Db    191 LMMVNPALPNAYAV----ASGDAKALTFFGF-IPVVGYQGTVLPAFFVGMIGARLENWLH 245

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120

Db    246 KRVPEALDLLITPFLTFLVMSILGLFAIGPVFHSVETVVLAATEWILALPFGIAGIIIGG 305

Qy    121 VYSPIVITGLHQSFPPIELELFNQGGSFIF-ATASMANIAQGAACLAVFFLAKSEKLKGL 179

Db    306 LQQVIVVTGVHHIFNFLETQLLAETKANPFNPLLSAATAGQVGAVLAVAVKTKSAKLKAL 365

Qy    180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGV 239

Db    366 AYPSALSAALGITEPAIFGVNLRYGKPFVMGLVGGSAGGFIAALVGLKATGMSVTVLPGL 425

Qy    240 VSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEA 299

Db    426 LLFLNSQMPMYIVSITVACAIAF--ALTY--YFGYADKEEDVSAKKPEAPAAAPVAETET 481

Qy    300 PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 359

Db    482 KSE------VIASPLDGEAVELSKVNDPVFSSEAMGKGIAVKPSGNTVYSPVNGTVQIAF 535

Qy    360 PSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAI 419

Db    536 ETGHAYGL--KSDNGA--EVLIHVGIDTVSMNGTGFDQKVAANQTVKVGDVLGTFDSAKI 591

Qy    420 KAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460

Db    592 AEAGLDDTTMVIITNTADYSEVKPLAAGQLAHGADLLELNK 632


RESULT  10
C95220
trehalose PTS system, IIABC components [imported] - Streptococcus pneumoniae (strain TIG
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: C95220
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: C95220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75956.1; PID:g14973388; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1884
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase


  Query Match             28.5%;  Score 673;  DB 2;  Length 655;
  Best Local Similarity   35.3%;  Pred. No. 5.7e-39;
  Matches  170;  Conservative   87;  Mismatches  173;  Indels   52;  Gaps   16;

Qy      1 MAMVFPSLVNGYDVAATMAAGEMP--MWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKF 58

Db    195 ICLVSPQLLNAYAVASTPAADIAANWVWNFGYFTVNRIGYQAQVIPALLAGLSLSYLEIF 254

Qy     59 LHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFG--GPVG-- 114

Db    255 WHKHIPEVISMIFVPFLSLIPALILAHTVLGP----IGWTIGQGLSSVVLAGLTGPVKWL 310

Qy    115 -GLLFGLVYSPIVITGLHQSFPPIELELF-NQGGSFIFATASMANIAQGAACLAVFFLAK 172

Db    311 FGAIFGALYAPFVITGLHHMTNAIDTQLIADAGGTALWPMIALSNIAQGSAVFAYYFMHR 370

Qy    173 -SEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVAL 231

Db    371 HDEREAQVSLPATISAYLGVTEPALFGVNVKYIYPFVAGMTGSALAGMLSVTFNVTAASI 430

Qy    232 GAAGFLGVVSIDAPDMVMF----LVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAP 287

Db    431 GIGGLPGILSIQPQYMLPFAGTMLVAIVVPMLLTF---------FFRKAGLFTK------ 475

Qy    288 VPAGTTKAEAE----APAEFSNDSTI------IQAPLTGEAIALSSVSDAMFASGKLGSG 337

Db    476 -TEGDTNLQAEFVAQEEAEFVNHEPVELTSVEIISPLTGQVKELSQATDPIFASGVMGQG 534

Qy    338 VAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNP 397

Db    535 LVIEPSQGELTSPVNGTVTVLFPTKHAIGI--VSDEG--VELLIHIGMDTVGLDGKGFES 590

Qy    398 LKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGAN 454

Db    591 LVVQGDHVTVGQQLIRFDMDVIKAAGLVTETPVIITNQDAYTATIP-GTYPT-TIQAGAS 648

Qy    455 LL 456

Db    649 LM 650


RESULT  11
AD1078
PTS system, beta-glucosides specific enzyme IIABC homolog lmo0027 [imported] - Lister
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AD1078
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AD1078
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-634 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98242.1; PID:g16409386; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0027
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera


  Query Match             28.2%;  Score 667;  DB 2;  Length 634;
  Best Local Similarity   34.6%;  Pred. No. 1.4e-38;
  Matches  156;  Conservative   81;  Mismatches  172;  Indels   42;  Gaps   10;

Qy      2 AMVFPSL--VNGYDVAATMAAG---EMPMWSLF-GLDVAQAGYQGTVLPVLVVSWILATI 55

Db    182 ALVYPTLAGITAGDPLYTLFAGTIFESPIHVTFLGIPVILMSYASSVIPIIVATYFGSKV 241

Qy     56 EKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGG 115

Db    242 EKGFKKIIPDVIKTFVVPFCTLLIVVPITFIVIGPIATWAGQLLGAGTIWVYNLSPIIAG 301

Qy    116 LLFGLVYSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSE 174

Db    302 LILGGFWQVFVIFGLHWGLVPVAINNLTVLGHDPILAMTFGASFAQIGAVLAVFFKSRNK 361

Qy    175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234

Db    362 KIKSLSIPAFISGIFGVTEPAIYGVTLPLKKPFIMSCIAGGIGGGIIGFAGSQTYIMGGL 421

Qy    235 GFLGVVSIDAP------DMVMFLVCAVVTFFIAFGAAIAYGL-----YLVRRNGSIDPDA 283

Db    422 GIFGLPNFFKPGSGISGEFWWVVIAIVISFILGFILTYVVGFKDPADVVVEQSNTVE--- 478
```

```
Qy    284 TAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPT 343
          | |  | |           | ||: || : |: | |  |:|| || ||||:||
Db    479 -------GETLIERET----------IPAPVVGEIVTLADVKDEAFSSGALGKGVAIIPT 521

Qy    344 KGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGD 403
          |::|:| :| :   ||:|||  : ||  ||:  ::|:||| ||| | |  |    ||||
Db    522 VGRVVAPAAGTVTTIFPTGHAIGITTK--DGA--EVLIHIGMDTVQLEGKFFTAHVKQGD 577

Qy    404 EVKAGELLCEFDIDAIKAAGYEVTTPIVVSN 434
          :: |:|| |||: ||||||:||||:||:|
Db    578 VIEKGQLLTEFDIEGIKAAGYDVTTPVVVTN 608


RESULT  12
F95200
PTS system IIABC components [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: F95200
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: F95200
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-627 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75799.1; PID:g14973217; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1722
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase


  Query Match             28.1%;  Score 664;  DB 2;  Length 627;
  Best Local Similarity   34.3%;  Pred. No. 2.3e-38;
  Matches  159;  Conservative   92;  Mismatches  187;  Indels   26;  Gaps   10;

Qy      1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          | :|  || | : ||     ||:   : ||  :   | ||:||  ::  : |  || :
Db    182 MMLVSGSLPNAWAVA---QGGEVTAMNFFGF-IPVVGLQGSVLPAFIIGVVGAKFEKAVR 237

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          | :    | |:|| :|||:   |    |||    | : :    : :      :|| | |
Db    238 KVVPDVIDLLVTPFVTLLVMSILGLFVIGPVFHVVENYILIATKAILSMPFGLGGFLIGG 297

Qy    121 VYSPIVITGLHQSFPPIELELFNQGGSFIF-ATASMANIAQGAACLAVFFLAKSEKLKGL 179
          |:  ||::|:|  |  :|::|     :  | |  : |  ||||| :||    |: ||| |
Db    298 VHQLIVVSGVHHIFNLLEVQLLAADHANPFNAIITAAMTAQGAATVAVGVKTKNPKLKTL 357

Qy    180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLG- 238
          |  : :|| |||||||||||||| | |||: :   |||| | ::  :     |     |
Db    358 AFPAALSAFLGITEPAIFGVNLRFRKPFFLSLIAGAIGGGLASILGLAGTGNGITIIPGT 417

Qy    239 VVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAE 298
          :: :    :  :|:   |:| : |     :|          : |||||   |   : : |
Db    418 MLYVGNGQLPQYLLMVAVSFALGFALTYMFGY-------EDEVDATAAAKRAEVAEEKEE 470

Qy    299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVV 357
          |||   |:: :   |: |: :||: |:| :|:|| :| |: : |::| : :|   :: :
Db    471 VAPAALQNETLV--TPIVGDVVALADVNDPVFSSGAMGQGIVVKPSQGVVYAPADAEVSI 528

Qy    358 AFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDID 417
          |||:|||| ::|:   :  ::|:|:| |||::||  |     ||::|||||::|  || :
Db    529 AFPTGHAFGLKTR----NGAEVLIHVGIDTVSMNGDGFETKVAQGNKVKAGDVLGTFDSN 584

Qy    418 AIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV 458
          | ||| : || ::|:|   |    || |   | :  |  :: |
Db    585 KIAAAGLDDTTMVIVTNTGDYASVAPVAT---GSVAKGDAVIEV 625
```

```
RESULT  13
E98067
phosphotransferase system enzyme II (EC 2.7.1.69) scrA [imported] - Streptococcus pne
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C;Accession: E98067
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E98067
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-627 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAL00370.1; PID:g15459232; GSPDB:GN00174
C;Genetics:
A;Gene: scrA
C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera
C;Keywords: phosphotransferase


  Query Match             27.9%;  Score 660;  DB 2;  Length 627;
  Best Local Similarity   34.3%;  Pred. No. 4.3e-38;
  Matches  159;  Conservative   92;  Mismatches  187;  Indels   26;  Gaps   10;

Qy      1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          | :|  || | : ||     ||:   : ||  :   | ||:||  ::  : |  || :
Db    182 MMLVSGSLPNAWAVA---QGGEVTAMNFFGF-IPVVGLQGSVLPAFIIGVVGAKFEKAVR 237

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          | :    | |:|| :|||:   |    |||    | : :    : :      :|| | |
Db    238 KVVPDVIDLLVTPFVTLLVMSILGLFVIGPVFHVVENYILIATKAILSMPFGLGGFLIGG 297

Qy    121 VYSPIVITGLHQSFPPIELELFNQGGSFIF-ATASMANIAQGAACLAVFFLAKSEKLKGL 179
          |:  ||::|:|  |  :|::|     :  | |  : |  ||||| :||    |: ||| |
Db    298 VHQLIVVSGVHHIFNLLEVQLLAADHANPFNAIITAAMTAQGAATVAVGVKTKNPKLKTL 357

Qy    180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLG- 238
          |  : :|| |||||||||||||| | |||: :   |||| | ::  :     |     |
Db    358 AFPAALSAFLGITEPAIFGVNLRFRKPFFLSLIAGAIGGGLASILGLAGTGNGITIIPGT 417

Qy    239 VVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAE 298
          :: :    :  :|:   |:| : |     :|          : |||||   |   : : |
Db    418 MLYVGNGQLPQYLLMVAVSFALGFALTYMFGY-------EDEVDATAAAKQAEVAEEKEE 470

Qy    299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVV 357
          |||   |:: :   |: |: :||: |:| :|:|| :| |: : |::| : :|   :: :
Db    471 VAPAALQNETLV--TPIVGDVVALADVNDPVFSSGAMGQGIAVKPSQGVVYALADAEVSI 528

Qy    358 AFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDID 417
          |||:|||| ::|:   :  ::|:|:| |||::||  |     ||::|||||::|  || :
Db    529 AFPTGHAFGLKTR----NGAEVLIHVGIDTVSMNGDGFEAKVAQGNKVKAGDVLGTFDSN 584

Qy    418 AIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV 458
          | ||| : || ::|:|   |    || |   | :  |  :: |
Db    585 KIAAAGLDDTTMVIVTNTADYASVAPVAT---GSVSKGDAVIEV 625


RESULT  14
F82432
PTS system, sucrose-specific IIBC component VCA0653 [imported] - Vibrio cholerae (str
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: F82432
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
```

```
A;Reference number: A82035; MUID:20406833
A;Accession: F82432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <HEI>
A;Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0653
A;Map position: 2
C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II


  Query Match           27.9%;  Score 659;  DB 2;  Length 479;
  Best Local Similarity  49.3%;  Pred. No. 3.7e-38;
  Matches  133;  Conservative   46;  Mismatches   87;  Indels    4;  Gaps    1;

Qy      1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          | || | |:||:   :   :| :| |:: | :: : ||||:|||||| ::||| ||  |
Db    194 MLMVHPDLLNGWGFGSASVSGTVPTWNILGFEIEKVGYQGSVLPVLVSAYILAKIENGLR 253

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          | :    | |:||:| : :|||||  :||  | || :|   |  |||  | ||| |||
Db    254 KIVPSVVDNLLTPMLAIFITGFLTFTVVGPLTRDVGFMLGDALNWLYDSAGFVGGALFGF 313

Qy    121 VYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
          :|:| ||||:| ||  || :|       ||:|||  |:|:|||||| |||  : |  ||
Db    314 IYAPFVITGMHHSFIAIETQLLADIVTTGGTFIFPIAAMSNIAQGAAALAVGVMTKETKL 373

Qy    177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
          ||:|  |||:|:|||||||:|||||:||:||   |  ||:  | | |||:|| ||||||
Db    374 KGVAIPSGVTALLGITEPAMFGVNLKLRYPFIAAICGAALASAFITLFNVKAQALGAAGL 433

Qy    237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAI 266
           |::||:   :  :::   ::|  ||   :
Db    434 PGIISINPQQIGYYIMGMAISFVAAFALTV 463


RESULT  15
D90038
PTS system, sucrose-specific IIBC component [imported] - Staphylococcus aureus (strain N
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: D90038
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: D90038
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-480 <KUR>
A;Cross-references: GB:BA000018; PID:g13702328; PIDN:BAB43469.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: scrA
C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II


  Query Match           27.9%;  Score 659;  DB 2;  Length 480;
  Best Local Similarity  48.6%;  Pred. No. 3.7e-38;
  Matches  139;  Conservative   53;  Mismatches   82;  Indels   12;  Gaps    5;

Qy      1 MAMVFPSLVNGYDVAATMAAGE-MPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
          | :| |||:: ||    : ||: :| | :||| : | |||| |||:|| ::|||:||| |
Db    194 MILVHPSLMSAYDFPKAVEAGKAIPYWDVFGLHINQVGYQGQVLPMLVAAYILASIEKGL 253

Qy     60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
           | :    | |:||:|:: :| ||||  :||  | :|  |: ||  ||:||| :|||:||
Db    254 RKVIPTVLDNLLTPLLSIFITAFLTFSFVGPITRQLGYWLSDGLTWLYEFGGAIGGLIFG 313

Qy    120 LVYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAK-SE 174
          |:|:|||||:| ||  :|  |      ||||||  |:|:|:||| | :| ||: | ::
Db    314 LLYAPIVITGMHHSFIAVETTLIADATKTGGSFIFPIATMSNVAQGGAAIAAFFIIKQNK 373

Qy    175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
          ||||:| |:|:||:|||||||:|||||:||:||   |  : || | || | :||:||| |
Db    374 KLKGVASAAGISALLGITEPAMFGVNLKLRYPFIGAIVGSGIGSAYIAFFKVKAIALGTA 433

Qy    235 GFLGVVSIDA--PDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGS 278
          |  | :||:      : : |   ::| |    ||   | | :|  :
Db    434 GLPGFISINPVHAGWLHYFVGMTISFII----AITVTLILSKRKAN 475


Search completed: September 25, 2002, 06:02:01
Job time: 417 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2002, 06:01:04 ; Search time 25.03 Seconds
(without alignments)
723.961 Million cell updates/sec

Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MAMVFPSLVNGYDVAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1058.5	44.8	651	1	PTSA_PEDPE	P43470 pediococcus
2	953.5	40.4	664	1	PTSA_STRMU	P12655 streptococc
3	701	29.7	480	1	PTSB_STAXY	P51184 staphylococ
4	678	28.7	474	1	PTSB_PASMU	Q9cjz2 pasteurella
5	657	27.8	609	1	PTBA_BACSU	P40739 bacillus su
6	646	27.3	479	1	PTSB_VIBAL	P22825 vibrio algi
7	640	27.1	631	1	PTBA_ERWCH	P26207 erwinia chr
8	570	24.1	470	1	PTTB_BACSU	P39794 bacillus su
9	567.5	24.0	625	1	PTBA_ECOLI	P08722 escherichia
10	492	20.8	456	1	PTSB_SALTY	P08470 salmonella
11	489	20.7	456	1	PTSB_KLEPN	P27219 klebsiella
12	462.5	19.6	674	1	PTGA_CORGL	Q45298 corynebacte
13	406	17.2	459	1	SACX_BACSU	P15400 bacillus su
14	390	16.5	460	1	PTSB_BACSU	P05306 bacillus su
15	367	15.5	473	1	PTTB_ECOLI	P36672 escherichia
16	321.5	13.6	699	1	PTGA_BACSU	P20166 bacillus su
17	317.5	13.4	651	1	PTAA_KLEPN	P45604 klebsiella
18	299	12.7	726	1	PTGA_STRPN	P35595 streptococc
19	298.5	12.6	648	1	PTAA_ECOLI	P09323 escherichia
20	294.5	12.5	324	1	PTGA_BACST	P42015 bacillus st
21	287.5	12.2	631	1	YBFS_BACSU	P39816 bacillus su
22	270.5	11.4	485	1	PTDA_ECOLI	P24241 escherichia
23	256	10.8	189	1	PTGA_BORBU	Q44840 borrelia bu
24	255	10.8	168	1	YPQE_BACSU	P50829 bacillus su
25	254.5	10.8	634	1	LACY_STRTR	P23936 streptococc
26	250.5	10.6	474	1	YFEV_ECOLI	P77272 escherichia
27	249.5	10.6	154	1	PTGA_MYCCA	P45618 mycoplasma
28	246	10.4	168	1	PTGA_ECOLI	P08837 escherichia
29	244.5	10.3	482	1	PTSB_VIBCH	Q9kvd9 vibrio chol
30	244	10.3	168	1	PTGA_SALTY	P02908 salmonella
31	238.5	10.1	165	1	PTGA_HAEIN	P45338 haemophilus
32	233	9.9	627	1	LACY_LACDE	P22733 lactobacill
33	231.5	9.8	161	1	PTGA_BUCAI	Q9wxi7 buchnera ap
34	230.5	9.8	940	1	PTGA_MYCPN	P75569 mycoplasma
35	221	9.4	641	1	RAFP_PEDPE	P43466 pediococcus
36	204	8.6	908	1	PTGA_MYCGE	P47315 mycoplasma
37	165.5	7.0	639	1	LACY_LEULA	Q48624 leuconostoc
38	159	6.7	483	1	PTVB_ECOLI	P32154 escherichia
39	151.5	6.4	658	1	HRSA_ECOLI	P54745 escherichia
40	144.5	6.1	527	1	PTIB_BACSU	P54715 bacillus su
41	136.5	5.8	359	1	PTWC_ECOLI	P32672 escherichia
42	130	5.5	578	1	PTFB_RHOCA	P23387 rhodobacter
43	126.5	5.4	580	1	PTFB_XANCP	P23355 xanthomonas
44	123.5	5.2	632	1	PTMA_BUCAI	P57635 buchnera ap
45	122	5.2	694	1	PTFA_MYCPN	P75039 mycoplasma

ALIGNMENTS

RESULT 1
PTSA_PEDPE
ID PTSA_PEDPE STANDARD; PRT; 651 AA.
AC P43470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS system, sucrose-specific IIABC component (EIIABC-SCR) (Sucrose-
DE permease IIABC component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Pediococcus pentosaceus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Pediococcus.
OX NCBI_TaxID=1255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PPE1.0;
RA Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXX-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC ----------

CC ----------
DR EMBL; Z32771; CAA83668.1; -.
DR EMBL; L32093; AAA25567.1; -.
DR HSSP; P20166; 1GPR.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;

```
KW   Phosphorylation; Transmembrane.
FT   DOMAIN        1     40       EIIB DOMAIN.
FT   DOMAIN        ?      ?       EIIC DOMAIN.
FT   DOMAIN      510    651       EIIA DOMAIN.
FT   MOD_RES      25     25       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     324    324       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     562    562       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   651 AA;  68454 MW;  C87BA09D550A77F8 CRC64;
```

Query Match 44.8%; Score 1058.5; DB 1; Length 651;
Best Local Similarity 46.7%; Pred. No. 1.4e-62;
Matches 221; Conservative 78; Mismatches 151; Indels 23; Gaps 5;

```
Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
Db     194 MIMVLPSLVNGYSVATTMAAGKMVYWNVFGLHVAQAGYQGQVLPVLGVAFILATLEKFFH 253

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
Db     254 KHIKGAFDFTFTPMFAIVITGFLTFTIVGPVLRTVSDALTNGLVGLYNSTGWIGMGIFGL 313

Qy     121 VYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
Db     314 LYSAIVITGLHQTFPAIETQLLANVAKTGGSFIFPVASMANIGQGAATLAIFFATKSQKQ 373

Qy     177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
Db     374 KALTSSAGVSALLGITEPAIFGVNLKMKFPFVFAAIASGIASAFLGLFHVLSVAMGPASV 433

Qy     237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAE 296
Db     434 IGFISIASKSIPAFMLSAVISFVVAFIPTFIY----AKRTLGDDRDQVKSPAPTSTV--- 486

Qy     297 AEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIV 356
Db     487 ------INVNDEIISAPVTGASESLKQVNDQVFSAEIMGKGAAIVPSSDQVVAPADGVIT 540

Qy     357 VAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDI 416
Db     541 VTYDSHHAYGIKTTA----GAEILIHLGLDTVNLNGEHFTTNVQKGDTVHQGDLLGTFDI 596

Qy     417 DAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANL--LNVAKKEAVPAT 467
Db     597 AALKAANYDPTVMLIVTNTANYANVERLKVTNVQAGEQLVALTAPAASSVAAT 649
```

```
RESULT   2
PTSA_STRMU
ID   PTSA_STRMU     STANDARD;      PRT;   664 AA.
AC   P12655;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PTS system, sucrose-specific IIABC component (EIIABC-SCR) (Sucrose-
DE   permease IIABC component) (Phosphotransferase enzyme II, ABC
DE   component) (EC 2.7.1.69) (EII-SCR).
GN   SCRA.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GS-5;
RX   MEDLINE=89123027; PubMed=2536656;
RA   Sato Y., Poy F., Jacobson G.R., Kuramitsu H.K.;
RT   "Characterization and sequence analysis of the scrA gene encoding
RT   enzyme IIScr of the Streptococcus mutans
RT   phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
RL   J. Bacteriol. 171:263-271(1989).
RN   [2]
RP   SEQUENCE OF 639-664 FROM N.A.
RC   STRAIN=GS-5;
RX   MEDLINE=93329360; PubMed=8336109;
RA   Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RT   "Isolation, characterization and sequence analysis of the scrK gene
RT   encoding fructokinase of Streptococcus mutans.";
RL   J. Gen. Microbiol. 139:921-927(1993).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22711; AAA26971.1; -.
DR   EMBL; D13175; BAA02466.1; -.
DR   PIR; B32243; B32243.
DR   HSSP; P08837; 1GLC.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Phosphorylation; Transmembrane.
FT   DOMAIN        1     40       EIIB DOMAIN.
FT   DOMAIN        ?      ?       EIIC DOMAIN.
FT   DOMAIN      533    664       EIIA DOMAIN.
FT   MOD_RES      25     25       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     331    331       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     585    585       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   664 AA;  69988 MW;  809E63E32281A9A1 CRC64;
```

Query Match 40.4%; Score 953.5; DB 1; Length 664;
Best Local Similarity 42.6%; Pred. No. 1.1e-55;
Matches 205; Conservative 89; Mismatches 156; Indels 31; Gaps 9;

```
Qy       1 MAMVFPSLVNGYDVAATMAAGEMP-------MWSLFGLDVAQAGYQGTVLPVLVVSWILA 53
Db     194 MIMVAPGAANIIGLAANAPISKAATIGAYTGFWNIFGLHVTQASYTYQVIPVLVAVWLLS 253

Qy      54 TIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPV 113
Db     254 ILEKFFHKRLPSAVDFTFTPLLSVIITGFLTFIVIGPVMKEVSDWLTNGIVWLYDTTGFL 313

Qy     114 GGLLFGLVYSPIVITGLHQSFPPIELEL---FNQG---GSFIFATASMANIAQGAACLAV 167
Db     314 GMGVFGALYSPVVMTGLHQSFPAIETQLISAFQNGTGHGDFIFVTASMANVAQGAATFAI 373

Qy     168 FFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIK 227
Db     374 YFLTKDKKMKGLSSSSGVSALLGITEPALFGVNLKYRFPFFCALIGSASAAAIAGLLQVV 433
```

```
Qy     228 AVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAP 287
           ||:||:||||| :|| |  :  ::|| :::| |||     ||   `  :  ::|  |  |
Db     434 AVSLGSAGFLGFLSIKASSIPFYVVCELISFAIAFAVTYGYG-----KTKAVDVFAAEAA 488

Qy     288 VPAGTTKAEAEAPAEFSNDST-------IIQAPLTGEAIALSSVSDAMFASGKLGSGVAI 340
           |      : : | | : :: :         :: ||| |||: |:||:| :|:|  :| |:||
Db     489 VEEAIEEVQ-EIPEEAASAANKAQVTDEVLAAPLAGEAVELTSVNDPVFSSEAMGKGIAI 547

Qy     341 VPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKK 400
            |:   : :|| | : :|| :|||: :  |:::|:  :||:||| |||:: |  |    :
Db     548 KPSGNTVYAPVDGTVQIAFDTGHAYGI--KSDNGA--EILIHIGIDTVSMEGKGFEQKVQ 603

Qy     401 QGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTY-GLGEIEAGANLLNVA 459
              ::| |::|  || | |  || : ||  :|:|      | |    | :  | :|| |
Db     604 ADQKIKKGDVLGTFDSDKIAEAGLDNTTMFIVTNTADYASVETLASSGTVAVGDSLLEVK 663

Qy     460 K 460
           |
Db     664 K 664


RESULT   3
PTSB_STAXY
ID   PTSB_STAXY     STANDARD;      PRT;   480 AA.
AC   P51184;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SCRA.
OS   Staphylococcus xylosus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
OX   NCBI_TaxID=1288;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 20267 / C2A;
RX   MEDLINE=94049686; PubMed=8232209;
RA   Wagner E., Goetz F., Brueckner R.;
RT   "Cloning and characterization of the scrA gene encoding the sucrose-
RT   specific Enzyme II of the phosphotransferase system from
RT   Staphylococcus xylosus.";
RL   Mol. Gen. Genet. 241:33-41(1993).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69800; CAA49461.1; -.
DR   HSSP; P05053; 1IBA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Inner membrane; Phosphorylation.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    480       EIIC DOMAIN.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     325    325       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   480 AA;  51326 MW;  AB4E1D9785D84E47 CRC64;


  Query Match             29.7%;  Score 701;  DB 1;  Length 480;
  Best Local Similarity   51.4%;  Pred. No. 3.3e-39;
  Matches  146;  Conservative   49;  Mismatches   81;  Indels    8;  Gaps    4;

Qy       1 MAMVFPSLVNGYDVAATMAAG-EMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
           | :| | |:: ||    : || |:| |:||||:: | |||| |||:|| ::|||||||
Db     194 MILVHPELMSAYDYPKALEAGKEIPHWNLFGLEINQVGYQGQVLPMLVATYILATIEKGL 253

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
           | :    | |:||:| :| |||:||  :||  | :|  |: ||  ||:||| :|||:||
Db     254 RKVIPTVLDNLLTPLLAILSTGFITFSFVGPLTRTLGYWLSDGLTWLYEFGGAIGGLIFG 313

Qy     120 LVYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAK-SE 174
           |:|:|||||||:| ||  || :|     : ||||||  |:|:|||||||| || ||: | ::
Db     314 LLYAPIVITGMHHSFIAIETQLIADSSSTGGSFIFPIATMSNIAQGAAALAAFFIIKENK 373

Qy     175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
           |||||:| |:|||||:|||||||:|||||:|||:||   |  : || | || | :|||:||| |
Db     374 KLKGVASAAGVSALLGITEPAMFGVNLKLRYPFIGAIVGSGIGSAYIAFFKVKAIALGTA 433

Qy     235 GFLGVVSIDAPD--MVMFLVCAVVTFFIAFGAAIAYGLYLVRRN 276
           |  | :||   :   : : :  :: | :|||   |       ||
Db     434 GIPGFISISGQNNGWLHYGIAMIIAFIVAFGVTYALSYRKKYRN 477


RESULT   4
PTSB_PASMU
ID   PTSB_PASMU     STANDARD;      PRT;   474 AA.
AC   Q9CJZ2;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SCRA OR PTSB OR PM1846.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (By similarity).
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE006222; AAK03930.1; -.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; FALSE_NEG.
KW   Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW   Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN      109    474       EIIC DOMAIN.
FT   TRANSMEM    109    129       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    182    202       POTENTIAL.
FT   TRANSMEM    229    249       POTENTIAL.
FT   TRANSMEM    264    284       POTENTIAL.
FT   TRANSMEM    303    323       POTENTIAL.
FT   TRANSMEM    345    365       POTENTIAL.
FT   TRANSMEM    376    396       POTENTIAL.
FT   TRANSMEM    403    423       POTENTIAL.
FT   TRANSMEM    444    464       POTENTIAL.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     324    324       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   474 AA;  49908 MW;  D26FA1A059603C01 CRC64;


  Query Match             28.7%;  Score 678;  DB 1;  Length 474;
  Best Local Similarity   49.5%;  Pred. No. 1.1e-37;
  Matches  135;  Conservative   51;  Mismatches   83;  Indels    4;  Gaps    1;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           | :| |:| :|:: | |:| | :  | :|||:: : |||||:|||| ||:|||:|| |
Db     194 MLLVHPALSDGWNYALTLAKGNIQYWHIFGLEIERVGYQGTVIPVLVASWVLATLEKNLR 253

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           | :    | ||||:  | :|| | |  ||| |  | ::: ||  |||  | ||| :||
Db     254 KVVPSFLDNLITPLFALFITGLLAFTVIGPIGREAGSLISTGLTWLYDTLGFVGGAIFGT 313

Qy     121 VYSPIVITGLHQSFPPIELELFNQ----GGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
           :|:|||||:||:|  :| :|  :    ||:|||  |:|:|||||||||   :: |  |:
Db     314 LYAPIVITGMHQTFIAVETQLLAEVARTGGTFIFPIAAMSNIAQGAACLGAAYVMKDAKV 373

Qy     177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
           :|:|  ||:||:|||||||:|||||  |:||   :  | |  |:|||||:||:||||||
Db     374 RGIAVPSGISALLGITEPAMFGVNLRYRYPFISAMIGAGISSAVIALFNVKAIALGAAGL 433

Qy     237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYG 269
             |: ||    : |: |  :::  |||   :  |
Db     434 PGIPSIKPDSLAMYCVGMLISASIAFTLTVILG 466


RESULT   5
PTBA_BACSU
ID   PTBA_BACSU     STANDARD;      PRT;   609 AA.
AC   P40739; Q45661;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   PTS system, beta-glucoside-specific IIABC component (EIIABC-BGL)
DE   (Beta-glucoside-permease IIABC component) (Phosphotransferase
DE   enzyme II, ABC component) (EC 2.7.1.69) (EII-BGL).
GN   BGLP OR N17C.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE=95189730; PubMed=7883710;
RA   Le Coq D., Lindner C., Krueger S., Steinmetz M., Stuelke J.;
RT   "New beta-glucoside (bgl) genes in Bacillus subtilis: the bglP gene
RT   product has both transport and regulatory functions similar to those
RT   of BglF, its Escherichia coli homolog.";
RL   J. Bacteriol. 177:1527-1535(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=95219088; PubMed=7704263;
RA   Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT   "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT   genome containing the hut and wapA loci.";
RL   Microbiology 141:337-343(1995).
RN   [3]
RP   SEQUENCE OF 1-182 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96204517; PubMed=8628237;
RA   Beloin C., Hirschbein L., le Hegarat F.;
RT   "Suppression of the Bgl+ phenotype of a delta hns strain of
RT   Escherichia coli by a Bacillus subtilis antiterminator binding
RT   site.";
RL   Mol. Gen. Genet. 250:761-766(1996).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z34526; CAA84286.1; -.
DR   EMBL; D31856; BAA06652.1; -.
DR   EMBL; D29985; BAA06256.1; -.
DR   EMBL; X85408; CAA59697.1; -.
DR   EMBL; Z99124; CAB15963.1; -.
DR   PIR; S47174; S47174.
DR   HSSP; P20166; 1GPR.
DR   SubtiList; BG10934; bglP.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Phosphorylation; Transmembrane; Complete proteome.
FT   DOMAIN        1     41       EIIB DOMAIN.
FT   DOMAIN        ?      ?       EIIC DOMAIN.
FT   DOMAIN      480    609       EIIA DOMAIN.
FT   MOD_RES      25     25       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     302    302       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     532    532       PHOSPHORYLATION (BY SIMILARITY).
FT   TRANSMEM    112    132       POTENTIAL.
```

```
FT   TRANSMEM    141    161       POTENTIAL.
FT   TRANSMEM    174    194       POTENTIAL.
FT   TRANSMEM    202    222       POTENTIAL.
FT   TRANSMEM    246    266       POTENTIAL.
FT   TRANSMEM    281    301       POTENTIAL.
FT   TRANSMEM    321    341       POTENTIAL.
FT   TRANSMEM    351    371       POTENTIAL.
FT   TRANSMEM    379    399       POTENTIAL.
FT   TRANSMEM    412    432       POTENTIAL.
FT   CONFLICT     75     75       A -> S (IN REF. 3).
FT   CONFLICT    288    288       L -> F (IN REF. 2).
FT   CONFLICT    434    434       E -> G (IN REF. 2).
FT   CONFLICT    436    436       A -> S (IN REF. 2).
FT   CONFLICT    449    450       DG -> HR (IN REF. 2).
FT   CONFLICT    549    549       I -> M (IN REF. 2).
FT   CONFLICT    552    552       G -> S (IN REF. 2).
SQ   SEQUENCE   609 AA;  64550 MW;  5F630C671D21FBED CRC64;


  Query Match             27.8%;  Score 657;  DB 1;  Length 609;
  Best Local Similarity   33.4%;  Pred. No. 3.4e-36;
  Matches  153;  Conservative   84;  Mismatches  181;  Indels   40;  Gaps    8;

Qy      13 DVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLIT 72
           |: | : ||: |: |  || |  | |  ||:|:|:  || : :||:: :    :   ::
Db     186 DLTALLGAGK-PI-SFIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASLKLIVV 243

Qy      73 PVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQ 132
           |  |||:   || | :||    :|: |: |:  |:|  | |  :|    :| |::||:|
Db     244 PTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMILLAGTFSLIIMTGMHY 303

Qy     133 SFPPIELELFNQGG-SFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGI 191
           :| || :    | |  ::    :||: |  |  |||  ::::| | ||  : ::|::||
Db     304 AFVPIMINNIAQNGHDYLLPAMFLANMGQAGASFAVFLRSRNKKFKSLALTTSITALMGI 363

Qy     192 TEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGA-AGFLGVVSIDAPDMVMF 250
           ||||::|||:||: ||   :   | |||   :  : :  :|  ||   :     |  :
Db     364 TEPAMYGVNMRLKKPFAAALIGGAAGGAFYGMTGVASYIVGGNAGLPSIPVFIGPTFIYA 423

Qy     251 LVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEAPAEFSNDSTII 310
           ::  |: |     ||   |   |  :||                    : ||       ||
Db     424 MIGLVIAFAAETAAAYLLGFEDVPSDGS------------------QQPAVHEGSREII 464

Qy     311 QAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAPPSGHAFAVRTK 370
            :|: ||  ||| | | :|::| :| | || | :|::|||| | :   | : ||  : :
Db     465 HSPIKGEVKALSEVKDGVFSAGVMGKGFAIEPEEGEVVSPVRGSVTTIFKTKHAIGITS- 523

Qy     371 AEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPI 430
                :||:||| ||| | |  |    |:||:|  |: |  ||:: |||||:| ||:
Db     524 ---DQGAEILIHIGLDTVKLEGQWFTAHIKEGDKVAPGDPLVSFDLEQIKAAGYDVITPV 580

Qy     431 VVSNYKK--TGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
           :|:|  :    ||   |           |  |||: |
Db     581 IVTNTDQYSFSPVKEIG----------KVQPKEALLA 607


RESULT   6
PTSB_VIBAL
ID   PTSB_VIBAL     STANDARD;      PRT;   479 AA.
AC   P22825;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SCRA.
OS   Vibrio alginolyticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=663;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91071601; PubMed=2174811;
RA   Blatch G.L., Scholle R.R., Woods D.R.;
RT   "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
RT   uptake-encoding region.";
RL   Gene 95:17-23(1990).
RN   [2]
RP   SEQUENCE OF 1-12 FROM N.A.
RX   MEDLINE=91285433; PubMed=2060795;
RA   Blatch G.L., Woods D.R.;
RT   "Nucleotide sequence and analysis of the Vibrio alginolyticus scr
RT   repressor-encoding gene (scrR).";
RL   Gene 101:45-50(1991).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76768; AAA27555.1; -.
DR   EMBL; M35009; AAA27557.2; -.
DR   PIR; JQ0781; JQ0781.
DR   HSSP; P05053; 1IBA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Inner membrane; Phosphorylation.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    479       EIIC DOMAIN.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     324    324       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   479 AA;  49890 MW;  FBF906B5170E3EB7 CRC64;


  Query Match             27.3%;  Score 646;  DB 1;  Length 479;
  Best Local Similarity   46.5%;  Pred. No. 1.4e-35;
  Matches  128;  Conservative   54;  Mismatches   87;  Indels    6;  Gaps    3;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           | || | |:||:       :| :|:|:: | :: : ||||:|||||| ::||| :|  |
Db     194 MLMVHPDLLNGWGFGGASVSGNIPVWNILGFEIQKVGYQGSVLPVLVSAFILAKVELGLR 253

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           | :    | |:||:| : : | |||  :||  | :| :|  ||  ||:  | ||| :|||
Db     254 KVIPSVLDNLLTPLLAIFIAGLLTFTVVGPFTRDIGFLLGDGLNWLYNTAGFVGGAVFGL 313

Qy     121 VYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
           :|:| ||||:| ||  || :|      ||:|||  |:|:|::|||| |||  ::| :|:
Db     314 IYAPFVITGMHHSFIAIETQLLADIATTGGTFIFPIAAMSNVSQGAAALAVGVMSKDKKM 373

Qy     177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
           ||:|  |||: :|||||||:|||||:||:||   :  ||:  | | :||:|| ||||||
Db     374 KGIAIPSGVTGLLGITEPAMFGVNLKLRYPFIAAVCAAALSSAFITMFNVKAQALGAAGL 433

Qy     237 LGVVSIDAPDMVMFLVCA-VVTFFIAFGAAIAYGL 270
           |::||  || : : :   |: |  ||   |  |:
```

```
Db    434 PGIISI-TPDKIGYYIAGMVIAFLTAFVLTIVLGI 467


RESULT   7
PTBA_ERWCH
ID   PTBA_ERWCH     STANDARD;      PRT;   631 AA.
AC   P26207;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   PTS system, beta-glucoside-specific IIABC component (EIIABC-BGL)
DE   (Beta-glucoside-permease IIABC component) (Phosphotransferase
DE   enzyme II, ABC component) (EC 2.7.1.69) (EII-BGL).
GN   ARBF.
OS   Erwinia chrysanthemi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=556;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92121114; PubMed=1732212;
RA   el Hassouni M., Henrissat B., Chippaux M., Barras F.;
RT   "Nucleotide sequences of the arb genes, which control beta-glucoside
RT   utilization in Erwinia chrysanthemi: comparison with the Escherichia
RT   coli bgl operon and evidence for a new beta-glycohydrolase family
RT   including enzymes from eubacteria, archeabacteria, and humans.";
RL   J. Bacteriol. 174:765-777(1992).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M81772; AAA24814.1; -.
DR   PIR; B42603; B42603.
DR   HSSP; P20166; 1GPR.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Phosphorylation; Transmembrane; Inner membrane.
FT   DOMAIN        1     41       EIIB DOMAIN.
FT   DOMAIN        ?      ?       EIIC DOMAIN.
FT   DOMAIN      501    631       EIIA DOMAIN.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     309    309       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     553    553       PHOSPHORYLATION (BY SIMILARITY).
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     99    119       POTENTIAL.
FT   TRANSMEM    120    140       POTENTIAL.
FT   TRANSMEM    146    166       POTENTIAL.
FT   TRANSMEM    175    195       POTENTIAL.
FT   TRANSMEM    206    226       POTENTIAL.
FT   TRANSMEM    248    268       POTENTIAL.
FT   TRANSMEM    272    292       POTENTIAL.
FT   TRANSMEM    295    315       POTENTIAL.
FT   TRANSMEM    328    348       POTENTIAL.
FT   TRANSMEM    358    378       POTENTIAL.
FT   TRANSMEM    385    405       POTENTIAL.
FT   TRANSMEM    407    427       POTENTIAL.
FT   TRANSMEM    434    454       POTENTIAL.
SQ   SEQUENCE   631 AA;  66984 MW;  7D0BD27A36BFFFDF CRC64;


  Query Match             27.1%;  Score 640;  DB 1;  Length 631;
  Best Local Similarity   33.8%;  Pred. No. 4.6e-35;
  Matches  154;  Conservative   86;  Mismatches  200;  Indels   16;  Gaps    8;

Qy       3 MVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKR 62
           :| ||::  ::    | | :       |: :    |  :|:|:|  ||:   :|| |::
Db     184 LVHPSMIAAFN---AMQAPDHSTLHFLGIPITFINYSSSVIPILFASWVSCKLEKPLNRW 240

Qy      63 LKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVY 122
           |        ||:| ::::  ||| |||:  |:  :|| | | ||     : | : | ::
Db     241 LHANIRNFFTPLLCIVISVPLTFLLIGPSATWLSQMLAGGYQWLYGLNSLLAGAVMGALW 300

Qy     123 SPIVITGLHQSFPPIELELFNQ-GGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAG 181
             || |||  | |: |  |:  |   :      | : |  | | |    :  | ||:||
Db     301 QVCVIFGLHWGFVPLMLNNFSVIGHDTLLPLLVPAVLGQAGATLGVLLRTQDLKRKGIAG 360

Qy     182 ASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVVS 241
           ::  :|: ||||||::|| | || ||  |    |:| |::   :   : |
Db     361 SAFSAAIFGITEPAVYGVTLPLRRPFIFGCIGGALGAAVMGYAHTTMYSFGFPSIFSFTQ 420

Qy     242 IDAPDMVMFLV-CAVVTFFIAFG-AAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEA 299
           :  |  |   |  ||:   :||  ||:    : | ::   :    ||  ||    || :|
Db     421 VIPPTGVDSSVWAAVIGTLLAFAFAALTSWSFGVPKD---ETQPAAADSPA--VLAETQA 475

Qy     300 PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 359
             |    | |:  :|| || : |  |:|  |||| :| |:|| ||:|:| :|| | :   |
Db     476 NAGAVRDETLF-SPLAGEVLLLEQVADRTFASGVMGKGIAIRPTQGRLYAPVDGTVASLF 534

Qy     360 PSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAI 419
           : ||  : ::        ::|:|:| ||| |:| :| |  : || |: |:|| |||  ||
Db     535 KTHHAIGLASR----GGAEVLIHVGIDTVRLDGRYFTPHVRVGDVVRQGDLLLEFDGPAI 590

Qy     420 KAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANL 455
           :|||||::|||||::| :    |     |:::| | |
Db     591 EAAGYDLTTPIVITNSEDYRGVEPVASGKVDANAPL 626


RESULT   8
PTTB_BACSU
ID   PTTB_BACSU     STANDARD;      PRT;   470 AA.
AC   P39794; O34771;
DT   01-FEB-1995 (Rel. 31, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable PTS system, trehalose-specific IIBC component (EIIBC-TRE)
DE   (Trehalose-permease IIBC component) (Phosphotransferase enzyme II, BC
DE   component) (EC 2.7.1.69)' (EII-TRE).
GN   TREP OR TREB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / MARBURG;
RX   MEDLINE=97074649; PubMed=8917076;
RA   Schoeck F., Dahl M.K.;
```

```
RT   "Analysis of DNA flanking the treA gene of Bacillus subtilis reveals
RT   genes encoding a putative specific enzyme IITre and a potential
RT   regulator of the trehalose operon.";
RL   Gene 175:59-63(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AC327;
RX   MEDLINE=97124190; PubMed=8969503;
RA   Yamamoto H., Uchiyama S., Sekiguchi J.;
RT   "Cloning and sequencing of a 40.6 kb segment in the 73 degrees-76
RT   degrees region of the Bacillus subtilis chromosome containing genes
RT   for trehalose metabolism and acetoin utilization.";
RL   Microbiology 142:3057-3065(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AC327;
RX   MEDLINE=97417488; PubMed=9272861;
RA   Yamamoto H., Uchiyama S., Nugroho F.A., Sekiguchi J.;
RT   "Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region
RT   of the Bacillus subtilis genome reveal genes for a new two-component
RT   system, three spore germination proteins, an iron uptake system and a
RT   general stress response protein.";
RL   Gene 194:191-199(1997).
RN   [4]
RP   SEQUENCE OF 324-470 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95379486; PubMed=7651129;
RA   Helfert C., Gotsche S., Dahl M.K.;
RT   "Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by
RT   a phospho-alpha-(1-1)-glucosidase encoded by the treA gene.";
RL   Mol. Microbiol. 16:111-120(1995).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z54245; CAA91014.1; -.
DR   EMBL; D83967; BAA23409.1; -.
DR   EMBL; D86417; BAA22289.1; -.
DR   EMBL; X80203; CAA56494.1; -.
DR   EMBL; Z99108; CAB12609.1; -.
DR   HSSP; P05053; 1IBA.
DR   SubtiList; BG11009; treP.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    470       EIIC DOMAIN.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   TRANSMEM    110    130       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
FT   TRANSMEM    183    203       POTENTIAL.
FT   TRANSMEM    234    254       POTENTIAL.
FT   TRANSMEM    263    283       POTENTIAL.
FT   TRANSMEM    301    321       POTENTIAL.
FT   TRANSMEM    326    346       POTENTIAL.
FT   TRANSMEM    347    367       POTENTIAL.
FT   TRANSMEM    375    395       POTENTIAL.
FT   TRANSMEM    403    423       POTENTIAL.
FT   TRANSMEM    443    463       POTENTIAL.
FT   CONFLICT    140    140       F -> S (IN REF. 1).
FT   CONFLICT    363    363       M -> L (IN REF. 1 AND 4).
FT   CONFLICT    465    465       A -> G (IN REF. 1 AND 4).
SQ   SEQUENCE   470 AA;  49999 MW;  7A741850A2697D53 CRC64;


  Query Match             24.1%;  Score 570;  DB 1;  Length 470;
  Best Local Similarity   42.8%;  Pred. No. 1.4e-30;
  Matches  116;  Conservative   54;  Mismatches   99;  Indels    2;  Gaps    1;

Qy       3 MVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKR 62
           :| | |:| :   |   :||:|:|:||||:| : |||| |||:|: |::|| || || ||
Db     197 LVHPDLLNAWGYGAAEQSGEIPVWNLFGLEVQKVGYQGQVLPILLASYMLAKIEVFLTKR 256

Qy      63 LKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVY 122
                 |:   :|||||| :|| |||    :|:|| || ::     :||||:|  |
Db     257 TPEGIQLLVVAPITLLLTGFASFIIIGPITFAIGNVLTSGLISVFGSFAALGGLLYGGFY 316

Qy     123 SPIVITGLHQSFPPIELELFNQ--GGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           | :||||:| :|  ::|:|     ||:|::   :::|||||:| ||: |: | || |||:
Db     317 SALVITGMHHTFLAVDLQLIGSKLGGTFLWPMLALSNIAQGSAALAMMFIVKDEKQKGLS 376

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
             ||:|| |||||||||||||| |:|| | : :: : |  |:   : | ::|  |  |:
Db     377 LTSGISAYLGITEPAIFGVNLRYRFPFIIAMVSSGLAGMYISSQGVLASSVGVGGVPGIF 436

Qy     241 SIDAPDMVMFLVCAVVTFFIAFGAAIAYGLY 271
           || :    | :   :   : |   ||  :
Db     437 SIMSQYWGAFAIGMAIVLIVPFAGTYAYARF 467


RESULT   9
PTBA_ECOLI
ID   PTBA_ECOLI     STANDARD;      PRT;   625 AA.
AC   P08722;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   PTS system, beta-glucoside-specific IIABC component (EIIABC-BGL)
DE   (Beta-glucoside-permease IIABC component) (Phosphotransferase
DE   enzyme II, ABC component) (EC 2.7.1.69) (EII-BGL).
GN   BGLF OR BGLC OR BGLS OR B3722.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=88009877; PubMed=3309161;
RA   Bramley H.F., Kornberg H.L.;
RT   "Nucleotide sequence of bglC, the gene specifying enzymeIIbgl of the
RT   PEP:sugar phosphotransferase system in Escherichia coli K12, and
RT   overexpression of the gene product.";
RL   J. Gen. Microbiol. 133:563-573(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=87222180; PubMed=3034860;
RA   Schnetz K., Toloczyki C., Rak B.;
RT   "Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide
RT   sequence, genetic organization, and possible evolutionary
RT   relationship to regulatory components of two Bacillus subtilis
RT   genes.";
RL   J. Bacteriol. 169:2579-2590(1987).
```

```
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=93315143; PubMed=7686882;
RA   Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT   "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT   genome: organizational symmetry around the origin of replication.";
RL   Genomics 16:551-561(1993).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON BGLG.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15746; AAA83837.1; -.
DR   EMBL; M16487; AAA23510.1; -.
DR   EMBL; L10328; AAA62073.1; -.
DR   EMBL; AE000449; AAC76745.1; -.
DR   PIR; C25977; C25977.
DR   PIR; A47616; A47616.
DR   HSSP; P20166; 1GPR.
DR   EcoGene; EG10115; bglF.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Phosphorylation; Transmembrane; Inner membrane; Complete proteome.
FT   DOMAIN        1     39       EIIB DOMAIN.
FT   DOMAIN        ?      ?       EIIC DOMAIN.
FT   DOMAIN      495    625       EIIA DOMAIN.
FT   MOD_RES      24     24       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     306    306       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     547    547       PHOSPHORYLATION (BY SIMILARITY).
FT   TRANSMEM     59     79       POTENTIAL.
FT   TRANSMEM    100    120       POTENTIAL.
FT   TRANSMEM    141    161       POTENTIAL.
FT   TRANSMEM    167    187       POTENTIAL.
FT   TRANSMEM    203    223       POTENTIAL.
FT   TRANSMEM    245    265       POTENTIAL.
FT   TRANSMEM    285    305       POTENTIAL.
FT   TRANSMEM    325    345       POTENTIAL.
FT   TRANSMEM    355    375       POTENTIAL.
FT   TRANSMEM    381    401       POTENTIAL.
FT   TRANSMEM    404    424       POTENTIAL.
FT   TRANSMEM    432    452       POTENTIAL.
FT   TRANSMEM    490    510       POTENTIAL.
FT   TRANSMEM    526    546       POTENTIAL.
SQ   SEQUENCE   625 AA;  66482 MW;  9E12A2207125C4E6 CRC64;
```

```
  Query Match             24.0%;  Score 567.5;  DB 1;  Length 625;
  Best Local Similarity   30.8%;  Pred. No. 2.7e-30;
  Matches  142;  Conservative   91;  Mismatches  211;  Indels   17;  Gaps    7;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHK 61
           |:| | ::  ::      |  :      |: |    |  :|:|::  :|: : :|: |:
Db     180 ALVHPLILTAFENGQKADALGL---DFLGIPVTLLNYSSSVIPIIFSAWLCSILERRLNA 236

Qy      62 RLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLV 121
            |       ||:| |::   :||: :||   |: :::| |   ||       | : |
Db     237 WLPSAIKNFFTPLLCLMVITPVTFLLVGPLSTWISELIAAGYLWLYQAVPAFAGAVMGGF 296

Qy     122 YSPIVITGLHQSFPPIELELFN-QGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           :   |: |||    |: :  |   |   :      | :||  | | ||   :  : | :|
Db     297 WQIFVMFGLHWGLVPLCINNFTVLGYDTMIPLLMPAIMAQVGAALGVFLCERDAQKKVVA 356

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
           |:: :::: ||||||::||||  ::|| |   : |:|  :|     |  : |        :
Db     357 GSAALTSLFGITEPAVYGVNLPRKYPFVIACISGALGATIIGYAQTKVYSFGLPSIFTFM 416

Qy     241 SIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPD-ATAAPVPAGTTKAEAEA 299
                 : | | |       |  || |   |   |::     ||   ||     |
Db     417 QTIPSTGIDFTVWASV-----IGGVIAIGCAFV---GTVMLHFITAKRQPAQGAPQEKTP 468

Qy     300 PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 359
                 :   | :|:||| : |  |:|  |||| || |:||:|: |:: |||:|:|   |
Db     469 EVITPPEQGGICSPMTGEIVPLIHVADTTFASGLLGKGIAILPSVGEVRSPVAGRIASLF 528

Qy     360 PSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAI 419
            : ||  :  :::||  |:||:|:| ||| |:|  |:     ||:|  |: |  ||| ||
Db     529 ATLHAIGI--ESDDG--VEILIHVGIDTVKLDGKFFSAHVNVGDKVNTGDRLISFDIPAI 584

Qy     420 KAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           : ||:::|||:::||      |  :|  :| ||  ||:: :
Db     585 REAGFDLTTPVLISNSDDFTDVLPHGTAQISAGEPLLSIIR 625
```

```
RESULT  10
PTSB_SALTY
ID   PTSB_SALTY     STANDARD;      PRT;   456 AA.
AC   P08470;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SCRA.
OS   Salmonella typhimurium.
OG   Plasmid pUR400.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88216186; PubMed=3285123;
RA   Ebner R., Lengeler J.W.;
RT   "DNA sequence of the gene scrA encoding the sucrose transport protein
RT   EnzymeII(Scr) of the phosphotransferase system from enteric bacteria:
RT   homology of the EnzymeII(Scr) and EnzymeII(Bgl) proteins.";
RL   Mol. Microbiol. 2:9-17(1988).
RN   [2]
RP   REVISIONS.
RX   MEDLINE=94018607; PubMed=8412665;
RA   Jahreis K., Lengeler J.W.;
RT   "Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid
RT   repressor for sucrose and D-fructose specific regulons from enteric
RT   bacteria.";
RL   Mol. Microbiol. 9:195-209(1993).
RN   [3]
RP   SEQUENCE OF 1-7 FROM N.A.
RC   STRAIN=6153-62;
```

```
RX   MEDLINE=91100329; PubMed=1846143;
RA   Hardesty C., Ferran C., Direnzo J.M.;
RT   "Plasmid-mediated sucrose metabolism in Escherichia coli:
RT   characterization of scrY, the structural gene for a
RT   phosphoenolpyruvate-dependent sucrose phosphotransferase system
RT   outer membrane porin.";
RL   J. Bacteriol. 173:449-456(1991).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
CC       COVALENTLY BOUND EIIA DOMAIN. INSTEAD, EII-SCR-MEDIATED
CC       PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
CC       GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67750; CAA47973.1; -.
DR   EMBL; Y00541; CAA68605.1; ALT_SEQ.
DR   EMBL; M38416; AAA98418.1; -.
DR   PIR; S01036; WQEBST.
DR   HSSP; P05053; 1IBA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW   Transmembrane; Inner membrane; Phosphorylation.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    456       EIIC DOMAIN.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     309    309       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   456 AA;  47864 MW;  98A6F1620AE50885 CRC64;


  Query Match             20.8%;  Score 492;  DB 1;  Length 456;
  Best Local Similarity   38.5%;  Pred. No. 1.8e-25;
  Matches  104;  Conservative   52;  Mismatches  100;  Indels   14;  Gaps    3;

Qy       3 MVFPSLVNGYDVAA---TMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
           :  |:| | : |||   ||        : || ::|  |||||| |||:  | :: :|| |
Db     186 LTHPALTNAWGVAAGFHTM--------NFFGFEIAMIGYQGTVFPVLLAVWFMSIVEKQL 237

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
             : :    | ::|| ||::::||:  : |||| | :|| ::  |  |    | : |||||
Db     238 RRAIPDALDLILTPFLTVIISGFIALLIIGPAGRALGDGISFVLSTLISHAGWLAGLLFG 297

Qy     120 LVYSPIVITGLHQSFPPIELELFNQ---GGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
           :|| |||||:| ||  :|  |      | :|:    :|||:||| |||||:|  |  |:
Db     298 GLYSVIVITGIHHSFHAVEAGLLGNPSIGVNFLLPIWAMANVAQGGACLAVWFKTKDAKI 357

Qy     177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
           | :   |  ||:||||| ||||:|||   ||   :   | ||| :   ::   |:|
Db     358 KAITLPSAFSAMLGITEAAIFGINLRFVKPFIAALIGGAAGGAWVVSVHVYMTAVGLTAI 417

Qy     237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAI 266
             |:  : |  :: :::  |: | :||  ::
Db     418 PGMAIVQASSLLNYIIGMVIAFGVAFTVSL 447


RESULT  11
PTSB_KLEPN
ID   PTSB_KLEPN     STANDARD;      PRT;   456 AA.
AC   P27219;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SCRA.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1033-5P14 / KAY2026;
RX   MEDLINE=91312133; PubMed=1649946;
RA   Schmid K., Ebner R., Jahreis K., Lengeler J.W., Titgemeyer F.;
RT   "A sugar-specific porin, ScrY, is involved in sucrose uptake in
RT   enteric bacteria.";
RL   Mol. Microbiol. 5:941-950(1991).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
CC       COVALENTLY BOUND EIIA DOMAIN. INSTEAD, EII-SCR-MEDIATED
CC       PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
CC       GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X57401; CAA40658.1; -.
DR   PIR; S15195; S15195.
DR   HSSP; P05053; 1IBA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Inner membrane; Phosphorylation.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    456       EIIC DOMAIN.
FT   TRANSMEM    112    132       POTENTIAL.
FT   TRANSMEM    144    164       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
FT   TRANSMEM    209    229       POTENTIAL.
FT   TRANSMEM    247    267       POTENTIAL.
FT   TRANSMEM    288    308       POTENTIAL.
FT   TRANSMEM    330    350       POTENTIAL.
FT   TRANSMEM    360    380       POTENTIAL.
FT   TRANSMEM    388    408       POTENTIAL.
FT   TRANSMEM    428    448       POTENTIAL.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
```

```
FT   MOD_RES     309    309       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   456 AA;  48022 MW;  4AFDF5405CAEFC66 CRC64;


  Query Match             20.7%;  Score 489;  DB 1;  Length 456;
  Best Local Similarity   38.5%;  Pred. No. 2.8e-25;
  Matches  104;  Conservative   51;  Mismatches  101;  Indels   14;  Gaps    3;

Qy       3 MVFPSLVNGYDVAA---TMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
           :  |:| | : |||   ||        : ||::||  |||||| |||:  | :: :|| |
Db     186 LTHPALTNAWGVAAGFHTM--------NFFGIEVAMIGYQGTVFPVLLAVWFMSMVEKRL 237

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
             : :    | ::|| ||::::||:  : |||| | :|| ::  |  |    | : |||||
Db     238 RRVIPDALDLILTPFLTVIISGFIALLLIGPAGRALGDGISFILSTLISHAGWLAGLLFG 297

Qy     120 LVYSPIVITGLHQSFPPIELELFNQ---GGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
           :|| |||||:| ||  ||  |      | :|:    :|||:||| || ||:|  |  |:
Db     298 GLYSVIVITGIHHSFHAIEAGLLGNPSIGVNFLLPIWAMANVAQGGACFAVWFKTKDAKI 357

Qy     177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
           | :   |  ||:||||| ||||:|||   ||   :   | ||| :   ::   |:|
Db     358 KAITLPSAFSAMLGITEAAIFGINLRFVKPFIAALVGGAAGGAWVVSMHVYMTAVGLTAI 417

Qy     237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAI 266
             |:  : |  :: :::   : | :||  ::
Db     418 PGMAIVQASSLLNYIIGMAIAFAVAFALSL 447


RESULT  12
PTGA_CORGL
ID   PTGA_CORGL     STANDARD;      PRT;   674 AA.
AC   Q45298;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PTS system, glucose-specific IIABC component (EIIABC-GLC) (Glucose-
DE   permease IIABC component) (Phosphotransferase enzyme II, ABC
DE   component) (EC 2.7.1.69) (EII-GLC/EIII-GLC).
GN   PTSG.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OG   Plasmid pBSBG2.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC   Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13869;
RA   Yoon K.-H.;
RT   "Cloning and nucleotide sequence of enzyme II of Brevibacterium
RT   lactofermentum phosphotransferase system.";
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L18875; AAA22992.1; -.
DR   HSSP; P08837; 1GLC.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Phosphorylation; Transmembrane; Plasmid.
FT   DOMAIN        1     43       EIIB DOMAIN.
FT   DOMAIN        ?      ?       EIIC DOMAIN.
FT   DOMAIN      542    674       EIIA DOMAIN.
FT   TRANSMEM    126    146       POTENTIAL.
FT   TRANSMEM    162    182       POTENTIAL.
FT   TRANSMEM    193    213       POTENTIAL.
FT   TRANSMEM    225    245       POTENTIAL.
FT   TRANSMEM    260    280       POTENTIAL.
FT   TRANSMEM    303    323       POTENTIAL.
FT   TRANSMEM    344    364       POTENTIAL.
FT   TRANSMEM    377    397       POTENTIAL.
FT   TRANSMEM    409    429       POTENTIAL.
FT   TRANSMEM    442    462       POTENTIAL.
FT   TRANSMEM    493    513       POTENTIAL.
FT   MOD_RES      28     28       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     594    594       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   674 AA;  71626 MW;  67A75AAF76E42FA2 CRC64;


  Query Match             19.6%;  Score 462.5;  DB 1;  Length 674;
  Best Local Similarity   30.0%;  Pred. No. 2.4e-23;
  Matches  139;  Conservative   79;  Mismatches  204;  Indels   41;  Gaps    9;

Qy      27 SLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFI 86
           ::||| :    | | | | |: :  |  :|| | | :      :  |  :||:    |
Db     218 TVFGLPMVLNDYSGQVFPPLIAAIGLYWVEKALKKIIPEAVQMVFVPFFSLLIMIPATAF 277

Qy      87 AIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQ-G 145
             :||    ||: ::  |: : :|   :  ::  |:|  :|  |||     | ::  |  |
Db     278 LLGPFGIGVGNGISSLLEAVNNFSPFILSIVIPLLYPFLVPLGLHWPLNAIMIQNLNTLG 337

Qy     146 GSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGAS--GVSAVL--GITEPAIFGVNL 201
             ||        | |        || :|  || : :   |  |: | |  ||:||:::|| |
Db     338 YDFIQGPMGAWNFACFGLVTGVFLIALKEKNRAMRQVSLGGMLAGLLGGISEPSLYGVLL 397

Qy     202 RLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIA 261
           | :  :|  :    :|| :: :|:||| |     |  :::| | |  :     :   |
Db     398 RFKKTYFRLLPGCLVGGIVMGIFDIKAYAF---VFTSLLTIPAMDPWLGYTVGIAAAF-- 452

Qy     262 FGAAIAYGLYLVRRNGSID---------------PDATAAPV----------PAGTTKA 295
           | : :    :  | :  |                | | ||||           || | |
Db     453 FTSMLLVLFFDYRSDAERDEAKAQMAAAEQTNNTPAAPAAPVAPAAGAAAAGGAAGAT-A 511

Qy     296 EAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKI 355
           |  |   :     | :|| | |: || | | :||:|||| |:|| ||   :|:|    :
Db     512 VATKPRLAAGQLVEITSPLEGHAVPLSEVPDPIFAAGKLGPGIAIEPTGNTVVAPADATV 571

Qy     356 VVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFD 415
           ::   |||| |:| :    | |::|:||| ||| | |  |    ::  :||||: |  ||
Db     572 ILVQKSGHAVALRLE----SGVELLIHIGLDTVQLGGEGFKVHVERKQQVKAGDPLITFD 627

Qy     416 IDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNV 458
             : |::    : ||:||||  | | :      : :|   :: |
Db     628 PEFIRSKNLPLITPVVVSNANKFGEIVGIEAAQADATTTVIKV 670


RESULT  13
```

```
SACX_BACSU
ID   SACX_BACSU     STANDARD;      PRT;   459 AA.
AC   P15400;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Negative regulatory protein of sacY.
GN   SACX OR SACS OR IPA-14R.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90337338; PubMed=2116367;
RA   Zukowski M.M., Miller L., Cogswell P., Chen K., Aymerich S.,
RA   Steinmetz M.;
RT   "Nucleotide sequence of the sacS locus of Bacillus subtilis reveals
RT   the presence of two regulatory genes.";
RL   Gene 90:153-155(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95020537; PubMed=7934828;
RA   Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA   Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA   Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA   Rapoport G., Danchin A.;
RT   "Bacillus subtilis genome project: cloning and sequencing of the 97
RT   kb region from 325 degrees to 333 degrees.";
RL   Mol. Microbiol. 10:371-384(1993).
RN   [3]
RP   SEQUENCE OF 348-459 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=92216127; PubMed=1806041;
RA   Glaser P., Kunst F., Debarbouille M., Vertes A., Danchin A.,
RA   Dedonder R.;
RT   "A gene encoding a tyrosine tRNA synthetase is located near sacS in
RT   Bacillus subtilis.";
RL   DNA Seq. 1:251-261(1991).
CC   -!- FUNCTION: Negatively regulates sacY.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29333; AAA75335.1; -.
DR   EMBL; X52480; CAA36719.1; -.
DR   EMBL; X73124; CAA51570.1; -.
DR   EMBL; Z99123; CAB15867.1; -.
DR   PIR; JU0293; JU0293.
DR   PIR; S16421; S16421.
DR   HSSP; P05053; 1IBA.
DR   SubtiList; BG10560; sacX.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Transcription regulation; Transferase; Phosphorylation;
KW   Complete proteome.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    459       EIIC DOMAIN.
FT   MOD_RES      25     25       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     136    136       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     433    433       PHOSPHORYLATION (BY SIMILARITY).

SQ   SEQUENCE   459 AA;  49024 MW;  A5C4E996ECDA3D40 CRC64;


  Query Match             17.2%;  Score 406;  DB 1;  Length 459;
  Best Local Similarity   37.9%;  Pred. No. 8.3e-20;
  Matches   96;  Conservative   47;  Mismatches   88;  Indels   22;  Gaps    6;

Qy      28 LFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIA 87
           ::|| :   ||||:::|:|:  :::: ||| |   :  : | :| | :|:::|| |  |
Db     206 IWGLHIPMMGYQGSMIPILLSVFVMSKIEKLLKSIVPKSLDVVIIPFITVMVTGCLALIV 265

Qy      88 IGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELF---NQ 144
           : ||   :| ::   :  :|| |   | ||| :|| ||::||| ||  ||  |    :
Db     266 MNPAASIIGQIMTQSIVYIYDHAGIAAGALFGGIYSTIVLSGLHHSFYAIEATLLANPHV 325

Qy     145 GGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLR 204
           | :|:    ||||:||| | ||||   |   || :|  : ::| ||| || :|||||:|
Db     326 GVNFLVPIWSMANVAQGGAGLAVFLKTKQSSLKKIALPASLTAFLGIVEPIVFGVNLKLI 385

Qy     205 WPFFIGIGTA---AIGGALIALFNIKAVALGAAGFLGVVSIDAP--------DMVMFLVC 253
            ||   || |   ||||| :    : | : |  | : ::||  |          |: ||:
Db     386 RPF---IGAAIGGAIGGAYVVAVQVVANSYGLTG-IPMISIVLPFGAANFVHYMIGFLIA 441

Qy     254 AV----VTFFIAF 262
           ||     | |: |
Db     442 AVSAFIATLFLGF 454


RESULT  14
PTSB_BACSU
ID   PTSB_BACSU     STANDARD;      PRT;   460 AA.
AC   P05306;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-SCR).
GN   SACP OR IPA-49D.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88097369; PubMed=3122206;
RA   Fouet A., Arnaud M., Klier A., Rapoport G.;
RT   "Bacillus subtilis sucrose-specific enzyme II of the
RT   phosphotransferase system: expression in Escherichia coli and
RT   homology to enzymes II from enteric bacteria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95020537; PubMed=7934828;
RA   Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA   Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA   Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA   Rapoport G., Danchin A.;
RT   "Bacillus subtilis genome project: cloning and sequencing of the 97
RT   kb region from 325 degrees to 333 degrees.";
RL   Mol. Microbiol. 10:371-384(1993).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC       AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC       PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC       PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC       THE SUGAR.
CC   -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
```

```
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   -!- SIMILARITY: 57% IDENTITY WITH THE NEGATIVE REGULATORY PROTEIN OF
CC       SACY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03006; AAA22727.1; -.
DR   EMBL; X73124; CAA51605.1; -.
DR   EMBL; Z99123; CAB15831.1; -.
DR   PIR; A39938; A39938.
DR   PIR; S39704; S39704.
DR   HSSP; P05053; 1IBA.
DR   SubtiList; BG10595; sacP.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Phosphorylation; Complete proteome.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    460       EIIC DOMAIN.
FT   MOD_RES      26     26       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     309    309       PHOSPHORYLATION (BY SIMILARITY).
FT   TRANSMEM    113    133       POTENTIAL.
FT   TRANSMEM    148    168       POTENTIAL.
FT   TRANSMEM    178    198       POTENTIAL.
FT   TRANSMEM    202    222       POTENTIAL.
FT   TRANSMEM    247    267       POTENTIAL.
FT   TRANSMEM    288    308       POTENTIAL.
FT   TRANSMEM    328    348       POTENTIAL.
FT   TRANSMEM    358    378       POTENTIAL.
FT   TRANSMEM    386    406       POTENTIAL.
FT   TRANSMEM    407    427       POTENTIAL.
FT   TRANSMEM    429    449       POTENTIAL.
SQ   SEQUENCE   460 AA;  49002 MW;  6A303DC042BFE379 CRC64;


  Query Match             16.5%;  Score 390;  DB 1;  Length 460;
  Best Local Similarity   44.6%;  Pred. No. 9.4e-19;
  Matches   91;  Conservative   34;  Mismatches   69;  Indels   10;  Gaps    4;

Qy      68 DFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVYSPIVI 127
           | |:|| :|:::|||: ||||||  | :|  :   |  :||  | | ||:||  || ||:
Db     246 DLLVTPFVTVIVTGFVAFIAIGPLGRALGSGITVALTYVYDHAGFVAGLIFGGTYSLIVL 305

Qy     128 TGLHQSFPPIELELF-NQGGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVS 186
           ||:| ||  ||  |  : | :::    ||||:||| | |||||:||  | | :|  :  |
Db     306 TGVHHSFHAIEAGLIADIGKNYLLPIWSMANVAQGGAGLAVFFMAKKAKTKEIALPAAFS 365

Qy     187 AVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVVSIDAP- 245
           | |||||| ||||||| | ||   :   |:||| :   :: | | |  | : :::| ||
Db     366 AFLGITEPVIFGVNLRYRKPFIAAMIGGALGGAYVVFTHVAANAYGLTG-IPMIAIAAPF 424

Qy     246 ---DMVMFL----VCAVVTFFIAF 262
               ::: :|    : ||  |  ||
Db     425 GFSNLIHYLIGMAIAAVSAFIAAF 448


RESULT  15
PTTB_ECOLI
ID   PTTB_ECOLI     STANDARD;      PRT;   473 AA.
AC   P36672;
DT   01-JUN-1994 (Rel. 29, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PTS system, trehalose-specific IIBC component (EIIBC-TRE) (Trehalose-
DE   permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE   (EC 2.7.1.69) (EII-TRE).
GN   TREB OR B4240.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=95332210; PubMed=7608078;
RA   Klein W., Horlacher R., Boos W.;
RT   "Molecular analysis of treB encoding the Escherichia coli enzyme II
RT   specific for trehalose.";
RL   J. Bacteriol. 177:4043-4052(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=95334362; PubMed=7610040;
RA   Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA   Blattner F.R.;
RT   "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT   region from 92.8 through 100 minutes.";
RL   Nucleic Acids Res. 23:2105-2119(1995).
RN   [3]
RP   REVISIONS TO 184-194.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC       SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC       -TRANSPORT SYSTEM. IT BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
CC       COVALENTLY BOUND EIIA DOMAIN. THE IIB DOMAIN CONTAINS THE
CC       PHOSPHORYL TRANSFER DOMAIN; THE IIC DOMAIN CONTAINS THE
CC       TRANSMEMBRANE CHANNEL. THE EII-TRE-MEDIATED PHOSPHORYLATION OF
CC       TREHALOSE REQUIRES THE ACTIVITY OF ENZYME IIA-GLC, A COMPONENT OF
CC       THE MAJOR GLUCOSE TRANSPORT SYSTEM.
CC   -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC       histidine + sugar phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U06195; AAC43381.1; -.
DR   EMBL; U14003; AAA97137.1; -.
DR   EMBL; AE000495; AAC77197.1; -.
DR   EcoGene; EG12127; treB.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Phosphotransferase system; Sugar transport; Transferase;
KW   Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
FT   DOMAIN        1      ?       EIIB DOMAIN.
FT   DOMAIN        ?    473       EIIC DOMAIN.
```

```
FT   MOD_RES      29     29       PHOSPHORYLATION (BY SIMILARITY).
FT   TRANSMEM    111    131       POTENTIAL.
FT   TRANSMEM    159    179       POTENTIAL.
FT   TRANSMEM    189    209       POTENTIAL.
FT   TRANSMEM    226    246       POTENTIAL.
FT   TRANSMEM    259    279       POTENTIAL.
FT   TRANSMEM    301    321       POTENTIAL.
FT   TRANSMEM    341    361       POTENTIAL.
FT   TRANSMEM    371    391       POTENTIAL.
FT   TRANSMEM    399    423       POTENTIAL.
FT   TRANSMEM    441    461       POTENTIAL.
FT   CONFLICT    126    127       GL -> PF (IN REF. 1).
FT   CONFLICT    184    194       GGTPILGIVLG -> AQRRSLVSCLA (IN REF. 2).
FT   CONFLICT    187    187       MISSING (IN REF. 1).
FT   CONFLICT    307    307       A -> Q (IN REF. 1).
FT   CONFLICT    429    473       PGILSIQPSYWQVFALAMAIAIIIPIVLTSFIYQRKYRLGT
FT                                LDIV -> RNSLDSTELLAGVCAGNGYRHHHPDCTHLVYLS
FT                                AEIPPGHAGHCLIFFGAQLRSHSQE (IN REF. 1).
SQ   SEQUENCE   473 AA;  51080 MW;  7437F8822B624944 CRC64;


  Query Match             15.5%;  Score 367;  DB 1;  Length 473;
  Best Local Similarity   30.1%;  Pred. No. 3.1e-17;
  Matches   80;  Conservative   63;  Mismatches  109;  Indels   14;  Gaps    5;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMP-MWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
           : :| | |:| |     :   ::| :|      :|: |||  |:| |:    |  ||  |
Db     195 VTLVSPQLMNAY-----LLGQQLPEVWDFGMFSIAKVGYQAQVIPALLAGLALGVIETRL 249

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLA----HGLQGLYDFGGPVGG 115
           : :      :: || :|:|  ||    |||  | :|| :|    | : | :    |:|
Db     250 KRIVPDYLYLVVVPVCSLILAVFLAHALIGPFGRMIGDGVAFAVRHLMTGSF---APIGA 306

Qy     116 LLFGLVYSPIVITGLHQSFPPIELELF-NQGGSFIFATASMANIAQGAACLAVFFLAKSE 174
            ||| :|:|:|||||:||:   |:|::  : ||: ::   :::|||||:| : :   ::
Db     307 ALFGFLYAPLVITGVHQTTLAIDLQMIQSMGGTPVWPLIALSNIAQGSAVIGIIISSRKH 366

Qy     175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
           : ::  : :|| ||:||||::|:||: |:|    :  : : | |  |  : |  :|
Db     367 NEREISVPAAISAWLGVTEPAMYGINLKYRFPMLCAMIGSGLAGLLCGLNGVMANGIGVG 426

Qy     235 GFLGVVSIDAPDMVMFLVCAVVTFFI 260
           |  |::||      :| :   :   |
Db     427 GLPGILSIQPSYWQVFALAMAIAIII 452


Search completed: September 25, 2002, 06:17:42
Job time: 998 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 25, 2002, 05:58:54 ; Search time 71.46 Seconds
(without alignments)
1132.964 Million cell updates/sec

Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MAMVFPSLVNGYDVAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	977	41.3	534	2	Q93ML1	Q93ml1 lactobacill
2	962	40.7	650	2	Q9S6S6	Q9s6s6 lactococcus
3	759.5	32.1	627	16	Q9L8G6	Q9l8g6 clostridium
4	695.5	29.4	630	16	Q9KG19	Q9kg19 bacillus ha
5	674	28.5	633	16	Q92FS7	Q92fs7 listeria in
6	673.5	28.5	636	16	Q9KF90	Q9kf90 bacillus ha
7	673	28.5	655	16	Q97NW9	Q97nw9 streptococc
8	664	28.1	627	16	Q97PB8	Q97pb8 streptococc
9	659	27.9	479	16	Q9KLT8	Q9klt8 vibrio chol
10	659	27.9	480	16	Q99RQ0	Q99rq0 staphylococ
11	658.5	27.9	644	2	Q9KJ80	Q9kj80 streptococc
12	647.5	27.4	674	16	Q99XQ4	Q99xq4 streptococc
13	632	26.7	620	16	Q99Y91	Q99y91 streptococc
14	612.5	25.9	621	2	Q48408	Q48408 klebsiella
15	612	25.9	628	16	Q97J79	Q97j79 clostridium
16	607	25.7	577	2	Q9L461	Q9l461 lactobacill
17	589.5	24.9	616	2	Q46129	Q46129 clostridium
18	588.5	24.9	636	16	Q9CFK9	Q9cfk9 lactococcus
19	570.5	24.1	612	16	Q97S38	Q97s38 streptococc
20	569.5	24.1	470	16	Q9KAS1	Q9kas1 bacillus ha
21	564.5	23.9	620	16	Q9A0X4	Q9a0x4 streptococc
22	530.5	22.5	618	2	Q9X565	Q9x565 enterococcu
23	517.5	21.9	480	2	Q9F8X3	Q9f8x3 pseudomonas
24	509	21.5	640	2	Q9RLU2	Q9rlu2 listeria mo
25	507	21.5	661	2	O68468	O68468 corynebacte
26	499	21.1	457	2	Q9F499	Q9f499 erwinia amy
27	496	21.0	475	16	Q99WC9	Q99wc9 staphylococ
28	487	20.6	458	16	Q9KBR9	Q9kbr9 bacillus ha
29	468.5	19.8	683	2	Q46072	Q46072 corynebacte
30	467	19.8	372	2	Q45622	Q45622 bacillus sp
31	462.5	19.6	681	2	Q93MZ4	Q93mz4 corynebacte
32	458.5	19.4	465	2	P94470	P94470 bacillus st
33	396.5	16.8	494	16	Q92CE9	Q92ce9 listeria in
34	388	16.4	451	2	Q9ZHJ9	Q9zhj9 clostridium
35	348.5	14.7	681	16	Q99X32	Q99x32 staphylococ
36	345	14.6	665	16	Q97LJ0	Q97lj0 clostridium
37	342.5	14.5	453	16	Q9K700	Q9k700 bacillus ha
38	341.5	14.5	692	2	Q53922	Q53922 staphylococ
39	339	14.3	478	16	Q9KTJ2	Q9ktj2 vibrio chol
40	323	13.7	484	16	Q99X29	Q99x29 staphylococ
41	314	13.3	195	16	Q9CGG4	Q9cgg4 lactococcus
42	307	13.0	652	2	Q938Q7	Q938q7 lactobacill
43	305.5	12.9	675	16	Q9KEK8	Q9kek8 bacillus ha
44	294	12.4	688	16	Q99R97	Q99r97 staphylococ
45	289.5	12.3	675	2	Q57071	Q57071 staphylococ

ALIGNMENTS

RESULT 1
Q93ML1
ID Q93ML1 PRELIMINARY; PRT; 534 AA.
AC Q93ML1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SUCROSE-SPECIFIC ENZYME II OF THE PTS (FRAGMENT).
GN SCRA.
OS Lactobacillus sakei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1599;
RN [1]
RP SEQUENCE FROM N.A.
RA Dudez A.-M., Chaillou S., Hissler L., Stentz R., Champomier-Verges M.,
RA Alpert C.-A., Zagorec M.;
RT "Physical and genetic map of the Lactobacillus sakei 23K chromosome.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF401046; AAK92528.1; -.
FT NON_TER 1 1
SQ SEQUENCE 534 AA; 56230 MW; 01804F9DE70C0089 CRC64;

Query Match 41.3%; Score 977; DB 2; Length 534;
Best Local Similarity 44.2%; Pred. No. 3.5e-53;
Matches 211; Conservative 90; Mismatches 138; Indels 38; Gaps 10;

```
Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           | :| |:||||| || ::| | |  | :|||::||||||| |:||: |::||| :||| |
Db      81 MMLVMPNLVNGYGVAESIATGHMTYWHVFGLNIAQAGYQGQVIPVIGVAFILANLEKFFH 140

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           | |    ||  ||:|::::|||||  :|||:| | : :   |   |   | ||  :|||
Db     141 KHLNDAVDFTFTPMLSIIITGFLTFTLVGPALRIVSNGVTDSLVWAYQTLGAVGMGIFGL 200

Qy     121 VYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
           || ||:|||||||| ||  |       ||||||  |:|||||||||  ||||: |:::
Db     201 GYSAIVLTGLHQSFPAIETTLLADIAKTGGSFIFPVAAMANIAQGAATFAVFFVTKNKQQ 260
```

```
Qy    177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
          | |  ::|:||:|||||||:|||||:|::|||||:  : |   :| | :: :|::| ||
Db    261 KSLTTSAGISAMLGITEPALFGVNLKLKFPFFIGLIASGISSFIIGLLHVLSVSMGPAGI 320

Qy    237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAE 296
          :| ::|    :  |:: |:::| |||     ||   ::                  ||: |
Db    321 IGFIAIAPKSIPSFMMGAIISFVIAFVGTYLYGKKAMK----------------TTEEE 363

Qy    297 A--EAPA-----EFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVS 349
             ||||     |   |   | ||:||  : |:|| | :|||  :| |:||:||   :::
Db    364 IINEAPATPEVVERLQDEK-ISAPVTGRIVDLASVPDPVFASEAMGKGIAIMPTSQDVLA 422

Qy    350 PVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGE 409
          ||:| | :|  :|||: :  |::||:  ::|:||| |||||||  |  : :||  |  |:
Db    423 PVTGVITIAANTGHAYGI--KSDDGA--EVLIHIGLDTVNLNGIGFEKIVQQGQHVSEGD 478

Qy    410 LLCEFDIDAIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV-AKKE 462
          ||  |||| || ||    | :|:|   | :  |: |     :: |  :: : |||:
Db    479 LLGHFDIDKIKQAGLTPLTMTIVTNTAGYAQVDPLLTVDKAAMQ-GEEIIQLHAKKD 534
```

RESULT 2
Q9S6S6
ID Q9S6S6 PRELIMINARY; PRT; 650 AA.
AC Q9S6S6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENZYME II SUCROSE PROTEIN (EC 2.7.1.69).
GN SACB.
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99173919; PubMed=10074089;
RA Luesink E.J., Marugg J.D., Kuipers O.P., De Vos W.M.;
RT "Characterization of the divergent sacBK and sacAR operons, involved
RT in sucrose utilization by Lactococcus lactis.";
RL J. Bacteriol. 181:1924-1926(1999).
DR EMBL; Z97015; CAB09690.1; -.
DR HSSP; P45618; 2GPR.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Transferase.
SQ SEQUENCE 650 AA; 69636 MW; 250E26F7664D4204 CRC64;

Query Match 40.7%; Score 962; DB 2; Length 650;
Best Local Similarity 42.9%; Pred. No. 3.8e-52;
Matches 203; Conservative 89; Mismatches 145; Indels 36; Gaps 7;

```
Qy      1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
          | || | |:|||:||  ::   |  | :||  |||||||| |||| |::||| :|:| |
Db    194 MIMVMPGLINGYNVAEAISNHTMTYWDIFGFKVAQAGYQGQVLPVIGVAFILAKLERFFH 253

Qy     61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
          | |    ||  ||:|::::|||||  :|||:|:| : |  || |||:  | :| |:||
Db    254 KYLNDAIDFTFTPLLSVIITGFLTFTIVGPALRFVSNGLTDGLVGLYNTLGALGMLVFGG 313

Qy    121 VYSPIVITGLHQSFPPIELELFNQ------GGSFIFATASMANIAQGAACLAVFFLAKSE 174
           || ||:|||||||| ||  |         || |||  |: ||:||  |  |: |: |:
Db    314 FYSAIVVTGLHQSFPAIETMLITNYQHSGIGGDFIFPVAACANMAQAGATFAILFVTKNI 373
```

```
Qy    175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
          | | ||  :||||:|||||||:||:||:|::|||| :| :|||   : ||:: ||:||:|
Db    374 KTKALAAPAGVSAILGITEPALFGINLKLKYPFFIALGASAIGSLFMGLFHVLAVSLGSA 433

Qy    235 GFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTK 294
          | :| :|| |   : |::   ::| |||     ||  :  :  ||  :           |
Db    434 GLIGFISIKAGYNLQFMISIFISFLIAFVVTSIYGRRMEAK--SITKE----------K 480

Qy    295 AEAEAPAEFSNDSTIIQ--------APLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346
          :  |  ::  :  ||          ||: |  | || ||| :|:   :| |:|| |  |:
Db    481 NKQNATTQYQPEKVIIDPVKSGELLAPINGFVIPLSDVSDPVFSKEIMGKGIAIKPKSGE 540

Qy    347 LVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVK 406
          | ||  |:|::|: :|||: ::||    :  ::|:||| |||::||  |    | | :||
Db    541 LFSPADGEIIIAYETGHAYGIKTK----NGGEVLLHIGIDTVSMNGNGFIQNVKVGQKVK 596

Qy    407 AGELLCEFDIDAIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLL 456
          ||:||  || : || :| : |  ||::|   | :  |::     :|: |  :|
Db    597 AGDLLGSFDKEEIKKSGLDDTVIIVITNSASYNEILPLSEN--VDIKVGEKIL 647
```

RESULT 3
Q9L8G6
ID Q9L8G6 PRELIMINARY; PRT; 627 AA.
AC Q9L8G6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SCRA (FUSION: PTS SYSTEM, BETA-GLUCOSIDES SPECIFIC IIABC
DE COMPONENT).
GN SCRA OR CAC0423.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=20391269; PubMed=10937490;
RA Tangney M., Mitchell W.J.;
RT "Analysis of a catabolic operon for sucrose transport and metabolism
RT in Clostridium acetobutylicum ATCC 824.";
RL J. Mol. Microbiol. Biotechnol. 2:71-80(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AF205034; AAF35839.1; -.
DR EMBL; AE007557; AAK78403.1; -.
DR HSSP; P20166; 1GPR.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 627 AA; 67596 MW; B835AB0238FCA436 CRC64;

Query Match 32.1%; Score 759.5; DB 16; Length 627;
Best Local Similarity 36.7%; Pred. No. 1.5e-39;

```
  Matches  163;  Conservative   80;  Mismatches  176;  Indels   25;  Gaps    5;

Qy       3 MVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHKR 62
           |: | | | :    |:  |     ::|||::   ||||||||:|:  |::: ||| | | |
Db     186 MIHPDLQNAW----TLGEGIKHTINIFGLNIGMVGYQGTVLPILISVWVMSYIEKGLRKI 241

Qy      63 LKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLVY 122
           :    | |:|| |||::|||   : |||  |:||| :: ||| ||:  |   |:||| :|
Db     242 VPEALDILLTPFLTLMITGFFAMVVIGPGGRFVGDEISLGLQTLYNTTGFFSGVLFGGLY 301

Qy     123 SPIVITGLHQSFPPIELELFNQGG---SFIFATASMANIAQGAACLAVFFLAKSEKLKGL 179
           | |||||:| ||  ||  |        :|:    ||||:||| | |||:|  : :|:| :
Db     302 SLIVITGIHHSFHAIEAGLLANPAIHKNFLLPIWSMANVAQGGAALAVYFKTRDKKMKSI 361

Qy     180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGV 239
           |  :  | :|||||||||||||||   ||  |    ||||  |    :   |:|  |  |:
Db     362 AAPASFSCLLGITEPAIFGVNLRYTKPFIAGALGGAIGGGYIVFTKVAMTAVGVTGIPGI 421

Qy     240 VSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEA 299
             :    : :::  :    :||| |    : |  :    : |      |        | |:
Db     422 AIVKQGSSFLNYIIAMI----LAFGGAFIIAMVLGIKEEITEEDLNKETVNKDIKVEEVES 477

Qy     300 PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 359
                   : :|: |: : | :| |  || | :| |: : |  |::|||: | :|  |
Db     478 ----------VVSPVNGKVVLLKNVPDKTFAEGLIGDGIGVDPEDGEVVSPIDGTVVHVF 527

Qy     360 PSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAI 419
           : || |:::|    : |::|:||| ||| : |  |      |:||| |: | :||:| :
Db     528 ETKHAIAMKSK----NGVEMLIHIGIDTVKMEGNGFKSFINDGEEVKKGDKLIQFDLDLV 583

Qy     420 KAAGYEVTTPIVVSNYKKTGPVNT 443
           |          :|:|::  | ||:
Db     584 KEKAVSPIVLTIVTNHEDMGFVNS 607


RESULT   4
Q9KG19
ID   Q9KG19      PRELIMINARY;      PRT;   630 AA.
AC   Q9KG19;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.
GN   BH0296.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001508; BAB04015.1; -.
DR   HSSP; P20166; 1AX3.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.
SQ   SEQUENCE   630 AA;  66978 MW;  6F0218011686ADD5 CRC64;


  Query Match           29.4%;  Score 695.5;  DB 16;  Length 630;
  Best Local Similarity   32.8%;  Pred. No. 1.5e-35;
  Matches  152;  Conservative  108;  Mismatches  182;  Indels   21;  Gaps    5;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHK 61
           |:::||:|  :| |  :         : ||: |    |  || |:|:  : :: :||| :|
Db     185 ALIYPSIVELHDSAIDV--------TFFGIPVVLMNYTSTVFPILLAVFAMSYVEKFCNK 236

Qy      62 RLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLV 121
           ::       :||:: |::   :| | :||   ::|: :|  :| :: |   : | :   :
Db     237 KIHEAVKNFVTPLILLVVIVPVTLIILGPIGVYLGNGIASVIQEIFTFSPVLAGAIVAGI 296

Qy     122 YSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           :  :|| |:|    || |  |  :|   | | |: |  :|  | | |    |::||| ||
Db     297 WQVLVIFGIHWGIIPIILNNLSVRGEDVIKAVAAPAVFSQAGAALGVMLRTKNKKLKALA 356

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
           |:: ::|: ||||||::|| | |: || : : :||:|||::  :   ||| || | | :
Db     357 GSTSITALFGITEPAVYGVTLPLKKPFIMAVISAAVGGAIVGHYGSVAVAPGAPGLLTIP 416

Qy     241 SI---DAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEA 297
                |    | |::  :::| :|     |   |:|     :| : | :       : :
Db     417 IFYPEDGRGFVAFVIAIIISFVLA-----AVLTYIVGFKDPVDDEDTLSNESGSENEVKR 471

Qy     298 EAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVV 357
           |   :  : |  |::|| || : |: | | :|:|| :| |||: | :|:||:|::| :
Db     472 EDDKKEPSASEEIKSPLKGEVVPLTEVQDHVFSSGAMGKGVAVRPKEGRLVAPINGTVTS 531

Qy     358 AFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDID 417
           | : ||  : :     :  :| :|:| ||| | | ||    :|||| ||::| |||::
Db     532 LFETKHAIGITS----DNGTEIFIHVGIDTVQLKGEHFTSFIEQGDEVAAGDVLLEFDVE 587

Qy     418 AIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           | ||||:| ||:::| |:    | |    |: :   |::| |
Db     588 RITAAGYDVITPVLITNAKQFSNVQTTDKREVTSEDLLIHVIK 630


RESULT   5
Q92FS7
ID   Q92FS7      PRELIMINARY;      PRT;   633 AA.
AC   Q92FS7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIN0026 PROTEIN.
GN   LIN0026.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596163; CAC95259.1; -.
DR   ListiList; LIN00026; -.
KW   Complete proteome.
SQ   SEQUENCE   633 AA;  67277 MW;  8AD80445D96C35A0 CRC64;
```

```
  Query Match             28.5%;  Score 674;  DB 16;  Length 633;
  Best Local Similarity   36.2%;  Pred. No. 3.3e-34;
  Matches  161;  Conservative   75;  Mismatches  179;  Indels   30;  Gaps   10;

Qy       2 AMVFPSL--VNGYDVAATMAAG---EMPMWSLF-GLDVAQAGYQGTVLPVLVVSWILATI 55
           |:|:|:|  ::  |   |: ||   | |:   | |: |    |  :|:|::: ::  : :
Db     181 ALVYPTLAGISAGDPIYTLFAGTIFESPIHVTFLGIPVILMSYASSVIPIILATYFGSKV 240

Qy      56 EKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGG 115
           ||   | :       : |  |||:   :||: |||   | | :|  |   :|:    | |
Db     241 EKGFKKIIPDVIKTFVVPFCTLLVVVPITFLVIGPIATWAGQLLGAGTIWVYNLSPVVAG 300

Qy     116 LLFGLVYSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSE 174
           |: |  :   || |||    |: :  |   |   : |    |: ||  | |||||  :::
Db     301 LILGGFWQVFVIFGLHWGLIPVAINNLTTLGSDPVLAMMFGASFAQIGAVLAVFFKTRNK 360

Qy     175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
           |:| |:  : :| : |:|||||:|| | |: || :     |:|| :|   : |   :|
Db     361 KIKSLSIPAFISGIFGVTEPAIYGVTLPLKKPFIMSCIAGAVGGGIIGFVSAKVYIMGGL 420

Qy     235 GFLGVVSIDAP-----DMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVP 289
           |  |: :   |          ::| |:|  ||         |  |    |  ||   |  |
Db     421 GIFGLPNFFQPGAGITSAFWWVVIAIVISFI-------LGFILTYVAGFKDP---AEAVV 470

Qy     290 AGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVS 349
           |   | |   |       | ||: || : |: | |  |:|| || ||||:|| |::|:
Db     471 EETNVTEGETLIE----RETIPAPVVGEIVTLADVKDEAFSSGALGKGVAIIPTVGRVVA 526

Qy     350 PVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGE 409
           | :| :   ||:|||  : |   ||:  ::|:||| ||| | |  |    |||| :: |:
Db     527 PAAGTVTTIFPTGHAIGITT--NDGA--EVLIHIGMDTVQLEGKFFTAHVKQGDVIEKGQ 582

Qy     410 LLCEFDIDAIKAAGYEVTTPIVVSN 434
           || ||||: ||||||:||||:||:|
Db     583 LLTEFDIEGIKAAGYDVTTPVVVTN 607


RESULT   6
Q9KF90
ID   Q9KF90      PRELIMINARY;      PRT;   636 AA.
AC   Q9KF90;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.
GN   BGLP OR BH0595.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001509; BAB04314.1; -.
DR   HSSP; P45618; 2GPR.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.

SQ   SEQUENCE   636 AA;  68437 MW;  434C0B12311716F2 CRC64;


  Query Match             28.5%;  Score 673.5;  DB 16;  Length 636;
  Best Local Similarity   34.0%;  Pred. No. 3.6e-34;
  Matches  163;  Conservative   78;  Mismatches  193;  Indels   45;  Gaps    9;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLF-------------GLDVAQAGYQGTVLPVLV 47
           ::|:|:||        :  || |:::||              |: |    |  :|:|:::
Db     183 SLVYPTLV-------VLTEGE-PLYTLFTGTIFESPVHITFLGIPVILMSYATSVIPIIL 234

Qy      48 VSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLY 107
           ::  : :|  | | :       : |  |||:   |||| |||   | | :|     :|
Db     235 AAYFASKVEARLRKIIPDVVKTFLVPFFTLLIVVPLTFIVIGPIATWAGQLLGQFTLWVY 294

Qy     108 DFGGPVGGLLFGLVYSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLA 166
           :    : |   | :   || |||    || :  |  ||   : |    |: ||  |  |
Db     295 NLSPIIAGAFLGGFWQVFVIFGLHWGLIPIAINNLVVQGSDPVLAMVFAASFAQIGAVAA 354

Qy     167 VFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNI 226
           |:   | :|:| |:  : :| : |:|||||:|| | |: || |    ||:|||:| ||
Db     355 VWLKIKQQKVKTLSVPAFISGIFGVTEPAIYGVTLPLKRPFIISCIAAAVGGAIIGLFRS 414

Qy     227 KAVALGAAGFLGVVSI----DAPDMVMF-LVCAVVTFFIAFGAAIAYGLYLVRRNGSIDP 281
           :   :|  |  |: |      |  |   : :| |||  |:  |  :  |    | | :
Db     415 QGYIIGGLGIFGIPSFLHPADGMDAGFWGIVIAVVVAFV-LGFILTYLFGLKSGNASDEQ 473

Qy     282 DATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIV 341
           | |    || : |             | :|  |  | || : |  |:|| || |:||
Db     474 TETKAHTSTGTGEKEE-----------ISSPFNGSVITLSEIKDEAFSSGALGEGIAIE 521

Qy     342 PTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQ 401
           |::|:| |||||  :   :|: ||  : |        ::|:||| ||| |:|  |     |
Db     522 PSEGKLFSPVSGMVTALYPTHHALGITT----DRGAELLIHIGLDTVQLDGKFFTAHTIQ 577

Qy     402 GDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           | :|: |:|| |||| ||||| ||||::|:|:|: | :       :: ||  || : :
Db     578 GAQVEKGDLLIEFDIKEIKAAGYAVTTPVIVTNHKQYGQLFLTDKQQVNAGDRLLELTR 636


RESULT   7
Q97NW9
ID   Q97NW9      PRELIMINARY;      PRT;   655 AA.
AC   Q97NW9;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TREHALOSE PTS SYSTEM, IIABC COMPONENTS.
GN   SP1884.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
DR   EMBL; AE007479; AAK75956.1; -.
DR   TIGR; SP1884; -.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
```

```
DR   InterPro; IPR001950; SUI1.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; UNKNOWN_1.
DR   PROSITE; PS01118; SUI1_1; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   655 AA;  69785 MW;  BD8C8C55C9DE0953 CRC64;


  Query Match             28.5%;  Score 673;  DB 16;  Length 655;
  Best Local Similarity   35.3%;  Pred. No. 4e-34;
  Matches  170;  Conservative   87;  Mismatches  173;  Indels   52;  Gaps   16;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMP--MWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKF 58
           : :| | |:| | ||:| ||      :|:     | : |||  |:| |:    |: :| |
Db     195 ICLVSPQLLNAYAVASTPAADIAANWVWNFGYFTVNRIGYQAQVIPALLAGLSLSYLEIF 254

Qy      59 LHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFG--GPVG-- 114
            || :      : | |:|:   |    :||   :|  :  ||  :   |  |||
Db     255 WHKHIPEVISMIFVPFLSLIPALILAHTVLGP----IGWTIGQGLSSVVLAGLTGPVKWL 310

Qy     115 -GLLFGLVYSPIVITGLHQSFPPIELELF-NQGGSFIFATASMANIAQGAACLAVFFLAK 172
            | :|| :|:| ||||||     |: :|  : ||: ::   :::||||||:|  | :|: :
Db     311 FGAIFGALYAPFVITGLHHMTNAIDTQLIADAGGTALWPMIALSNIAQGSAVFAYYFMHR 370

Qy     173 -SEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVAL 231
             |:   ::  : :|| ||:||||:||||::  :|| |:  :|: | |   ||: | ::
Db     371 HDEREAQVSLPATISAYLGVTEPALFGVNVKYIYPFVAGMTGSALAGMLSVTFNVTAASI 430

Qy     232 GAAGFLGVVSIDAPDMVMF----LVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAP 287
           |  |  |::||    |: |    ||  ||   : |           |: |
Db     431 GIGGLPGILSIQPQYMLPFAGTMLVAIVVPMLLTF---------FFRKAGLFTK------ 475

Qy     288 VPAGTTKAEAE----APAEFSNDSTI------IQAPLTGEAIALSSVSDAMFASGKLGSG 337
             | |  :||       ||| |   :       | :||||:   ||  :| :|||| :| |
Db     476 -TEGDTNLQAEFVAQEEAEFVNHEPVELTSVEIISPLTGQVKELSQATDPIFASGVMGQG 534

Qy     338 VAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNP 397
           : | |::|:| |||:| : | ||: ||  :   :::|  |::|:||| ||| |:|  |
Db     535 LVIEPSQGELTSPVNGTVTVLFPTKHAIGI--VSDEG--VELLIHIGMDTVGLDGKGFES 590

Qy     398 LKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGAN 454
           |  ||| |  |: |  ||:| |||||    ||::::|   |  | |  ||    |:|||:
Db     591 LVVQGDHVTVGQQLIRFDMDVIKAAGLVTETPVIITNQDAYTATIP-GTYPT-TIQAGAS 648

Qy     455 LL 456
           |:
Db     649 LM 650


RESULT   8
Q97PB8
ID   Q97PB8      PRELIMINARY;      PRT;   627 AA.
AC   Q97PB8;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PTS SYSTEM IIABC COMPONENTS.
GN   SP1722.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIGR4;
RX   MEDLINE=21357209; PubMed=11463916;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA   Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
DR   EMBL; AE007465; AAK75799.1; -.
DR   TIGR; SP1722; -.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.
SQ   SEQUENCE   627 AA;  65683 MW;  9298CDD7B2E6B85F CRC64;


  Query Match             28.1%;  Score 664;  DB 16;  Length 627;
  Best Local Similarity   34.3%;  Pred. No. 1.4e-33;
  Matches  159;  Conservative   92;  Mismatches  187;  Indels   26;  Gaps   10;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           | :|  || | : ||      ||:   : ||  :   | ||:|||  ::  : |  || :
Db     182 MMLVSGSLPNAWAVA---QGGEVTAMNFFGF-IPVVGLQGSVLPAFIIGVVGAKFEKAVR 237

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           | :    | |:|| :|||:   |    |||    | : :    : :       :|| | |
Db     238 KVVPDVIDLLVTPFVTLLVMSILGLFVIGPVFHVVENYILIATKAILSMPFGLGGFLIGG 297

Qy     121 VYSPIVITGLHQSFPPIELELFNQGGSFIF-ATASMANIAQGAACLAVFFLAKSEKLKGL 179
           |:  ||::|:|  |  :|::|     :  | |  : |  ||||| :||    |: ||| |
Db     298 VHQLIVVSGVHHIFNLLEVQLLAADHANPFNAIITAAMTAQGAATVAVGVKTKNPKLKTL 357

Qy     180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLG- 238
           |  : :|| |||||||||||||| | |||: :   ||||  | ::  :     |     |
Db     358 AFPAALSAFLGITEPAIFGVNLRFRKPFFLSLIAGAIGGGLASILGLAGTGNGITIIPGT 417

Qy     239 VVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAE 298
           :: :    :  :|:   |:| : |     :|              : |||||   |   : : |
Db     418 MLYVGNGQLPQYLLMVAVSFALGFALTYMFGY-------EDEVDATAAAKRAEVAEEKEE 470

Qy     299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVV 357
            |||   |:: :   |: |: :||: |:| :|:|| :| |: : |::| : :|   :: :
Db     471 VAPAALQNETLV--TPIVGDVVALADVNDPVFSSGAMGQGIVVKPSQGVVYAPADAEVSI 528

Qy     358 AFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDID 417
           |||:|||| ::|:    :  ::|:|:| |||::||  |     ||::||||::|  || :
Db     529 AFPTGHAFGLKTR----NGAEVLIHVGIDTVSMNGDGFETKVAQGNKVKAGDVLGTFDSN 584

Qy     418 AIKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV 458
           | ||| : || ::|:|   |    || |   | :  |  :: |
Db     585 KIAAAGLDDTTMVIVTNTGDYASVAPVAT---GSVAKGDAVIEV 625


RESULT   9
Q9KLT8
ID   Q9KLT8      PRELIMINARY;      PRT;   479 AA.
AC   Q9KLT8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN   VCA0653.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004395; AAF96554.1; -.
DR   TIGR; VCA0653; -.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.
SQ   SEQUENCE   479 AA;  50292 MW;  DC0E129229957E3C CRC64;


  Query Match             27.9%;  Score 659;  DB 16;  Length 479;
  Best Local Similarity   49.3%;  Pred. No. 2e-33;
  Matches  133;  Conservative   46;  Mismatches   87;  Indels    4;  Gaps    1;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           | || | |:||:   :   :| :| |:: | :: : ||||:|||||| ::||| ||  |
Db     194 MLMVHPDLLNGWGFGSASVSGTVPTWNILGFEIEKVGYQGSVLPVLVSAYILAKIENGLR 253

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           | :    | |:||:| : :||||||  :||  | || :|   |  |||  | ||| |||
Db     254 KIVPSVVDNLLTPMLAIFITGFLTFTVVGPLTRDVGFMLGDALNWLYDSAGFVGGALFGF 313

Qy     121 VYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAKSEKL 176
           :|:| ||||:| ||  || :|       ||:|||  |:|:|||||| |||  : |  ||
Db     314 IYAPFVITGMHHSFIAIETQLLADIVTTGGTFIFPIAAMSNIAQGAAALAVGVMTKETKL 373

Qy     177 KGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGF 236
           ||:|  |||:|:|||||||:|||||:||:||   |  ||:  | | |||:|| ||||||
Db     374 KGVAIPSGVTALLGITEPAMFGVNLKLRYPFIAAICGAALASAFITLFNVKAQALGAAGL 433

Qy     237 LGVVSIDAPDMVMFLVCAVVTFFIAFGAAI 266
           |::||:   :  :::   ::|  ||   :
Db     434 PGIISINPQQIGYYIMGMAISFVAAFALTV 463


RESULT  10
Q99RQ0
ID   Q99RQ0      PRELIMINARY;      PRT;   480 AA.
AC   Q99RQ0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN   SCRA OR SA2167 OR SAV2377.
OS   Staphylococcus aureus (strain N315), and
OS   Staphylococcus aureus (strain Mu50).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
OX   NCBI_TaxID=158879, 158878;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003137; BAB43469.1; -.
DR   EMBL; AP003365; BAB58539.1; -.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.
SQ   SEQUENCE   480 AA;  51232 MW;  D88607F6E0AF2E84 CRC64;


  Query Match             27.9%;  Score 659;  DB 16;  Length 480;
  Best Local Similarity   48.6%;  Pred. No. 2e-33;
  Matches  139;  Conservative   53;  Mismatches   82;  Indels   12;  Gaps    5;

Qy       1 MAMVFPSLVNGYDVAATMAAGE-MPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFL 59
           | :| |||:: ||    : ||: :| | :||| : | |||| |||:|| ::|||:||| |
Db     194 MILVHPSLMSAYDFPKAVEAGKAIPYWDVFGLHINQVGYQGQVLPMLVAAYILASIEKGL 253

Qy      60 HKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFG 119
           | :    | |:||:|:: :| ||||  :||  | :|  |: ||  ||:||| :|||:||
Db     254 RKVIPTVLDNLLTPLLSIFITAFLTFSFVGPITRQLGYWLSDGLTWLYEFGGAIGGLIFG 313

Qy     120 LVYSPIVITGLHQSFPPIELELF----NQGGSFIFATASMANIAQGAACLAVFFLAK-SE 174
           |:|:|||||||:| ||  :|  |       ||||||  |:|:|:||| | :| ||: | ::
Db     314 LLYAPIVITGMHHSFIAVETTLIADATKTGGSFIFPIATMSNVAQGGAAIAAFFIIKQNK 373

Qy     175 KLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAA 234
           ||||:| |:|:|||:|||||||:|||||:||:||   |  : || | || | :||:||| |
Db     374 KLKGVASAAGISALLGITEPAMFGVNLKLRYPFIGAIVGSGIGSAYIAFFKVKAIALGTA 433

Qy     235 GFLGVVSIDA--PDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGS 278
           |  | :||:      : : |   ::| |    ||   | | :|  :
Db     434 GLPGFISINPVHAGWLHYFVGMTISFII----AITVTLILSKRKAN 475


RESULT  11
Q9KJ80
ID   Q9KJ80      PRELIMINARY;      PRT;   644 AA.
AC   Q9KJ80;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BETA-GLUCOSIDE-SPECIFIC EII PERMEASE.
GN   BGLP.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NG8;
RX   MEDLINE=20340959; PubMed=10878120;
RA   Cote C.K., Cvitkovitch D., Bleiweis A.S., Honeyman A.L.;
RT   "A novel beta-glucoside-specific PTS locus from Streptococcus mutans
RT   that is not inhibited by glucose.";
RL   Microbiology 146:1555-1563(2000).
DR   EMBL; AF206272; AAF89975.1; -.
DR   HSSP; P20166; 1GPR.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
SQ   SEQUENCE   644 AA;  69282 MW;  723B7FBBDD2794EB CRC64;
```

```
  Query Match             27.9%;  Score 658.5;  DB 2;  Length 644;
  Best Local Similarity   33.8%;  Pred. No. 3.1e-33;
  Matches  160;  Conservative   89;  Mismatches  196;  Indels   29;  Gaps   12;

Qy       2 AMVFPSLVNGYDV---AATMAAGEM---PMWS-LFGLDV--AQAGYQGTVLPVLVVSWIL 52
Db     185 ALVYPNIAASFTVKHPLYTLFQGTLIVSPIYSTFFGIPIIFPASSYLQTVLPIVVAIWAG 244

Qy      53 ATIEKPLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGP 112
Db     245 SKIETFFKKIIPDVVKVFVVPFFTILITVPLSFLVIGPVMSWASDLVGAIFTGIYGFNPV 304

Qy     113 VGGLLFGLVYSPIVITGLHQSFPPIE-LELFNQGGSFIFATASMANIAQGAACLAVFFLA 171
Db     305 IYGIVLGAMWQVLVMFGLHWGLVPLAILELQKGPGVILVATIAIC-FAQAGSLLNIMMRT 363

Qy     172 KSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVAL 231
Db     364 NENKVRQLSIPAFISALFGVTEPAIYGITLPMRVPFIMTCVSGAISGAYLALFNVKMQVM 423

Qy     232 GAAGFLGVVSIDAPD----MVMFLVCAVVTFFIAFGAAIAYGL-YLVRRNGSIDPDATAA 286
Db     424 GGMGLFAIPSFIDPKNSMILIHFLIAIAMNFVLGFVLTQFIKIPYLYGEPTSTDSDVDDK 483

Qy     287 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346
Db     484 EEPVKELK-------EIKQE--IISSPLIGKVVKLENVPDEVFASGAMGKGIAIDPDDGI 534

Qy     347 LVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVK 406
Db     535 VVAPTKGEVTLVFPTKHAVGLRT--ENGA--EILIHIGMDTVSLAGKGFKSFVEVGDHVE 590

Qy     407 AGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
Db     591 AGQTLLEFDVNAIKAAGLPVITPVIVTNSDQFEDVLTTQERTVEAGNYLLTAVK 644


RESULT  12
Q99XQ4
ID   Q99XQ4      PRELIMINARY;      PRT;   674 AA.
AC   Q99XQ4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE PTS SYSTEM ENZYME II.
GN   SPY2097.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1314;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX   MEDLINE=21192684; PubMed=11296296;
RA   Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA   Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA   Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
RA   Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT   "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR   EMBL; AE006630; AAK34748.1; -.
DR   HSSP; P20166; 1GPR.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; UNKNOWN_1.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   674 AA;  72489 MW;  4BAA26B40E784878 CRC64;


  Query Match             27.4%;  Score 647.5;  DB 16;  Length 674;
  Best Local Similarity   34.4%;  Pred. No. 1.6e-32;
  Matches  155;  Conservative   86;  Mismatches  178;  Indels   31;  Gaps   11;

Qy       1 MAMVFPSLVNGYDVAATMAA--GEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKF 58
Db     212 ICLVSPQLLNAYAVAGTPAAEIAKNWVWDFGFFTINRIGYQAQVIPALLAGLSLAYLEIF 271

Qy      59 LHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFG--GPVG-- 114
Db     272 WRKRIPEVVSMIFVPFLSLIPALILAHTVLGP----IGWTIGKGISFVVLAGLTGPVKWL 327

Qy     115 -GLLFGLVYSPIVITGLHQSFPPIELELFNQGG---SFIFATASMANIAQGAACLAVFFL 170
Db     328 FGAIFGALYAPLVITGLHHMTNAIDTQLIADTATRTTGLWPMIALSNIAQGSAVFAYYLM 387

Qy     171 AKSEKLKG-LAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAV 229
Db     388 NRHEEREAEISLPAAISAYLGVTEPALFGVNVKYVYPFVAGMIGSGIAGLLSTTFNVQAN 447

Qy     230 ALGAAGFLGVVSIDAPDMVMFLVC----AVVTFFIAFGAAIAYGLYLVRRNGSIDPDATA 285
Db     448 SIGVGGLPGFMAINVKYMIPFFICMAVAIVVPMFLTFFFRKSHIMTKTEDEAKL-PETPV 506

Qy     286 APVPAGTTKAEAEAPAEFSNDSTI-IQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTK 344
Db     507 SDAPVAT------APHKTMQGTVITLTSPLTGEVKALSEAVDPVFAQGVMGQGALLQPTE 560

Qy     345 GQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDE 404
Db     561 GVLVAPCDAEVSVLFPTKHAICLVTT----EGLELLMHIGMDTVNLDGQGFEALVKQGDQ 616

Qy     405 VKAGELLCEFDIDAIKAAGYEVTTPIVVSN 434
Db     617 VKAGQTLIQFDIAAISEAGYATETPLVVTN 646


RESULT  13
Q99Y91
ID   Q99Y91      PRELIMINARY;      PRT;   620 AA.
AC   Q99Y91;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE SUCROSE-SPECIFIC PTS PERMEASE, ENZYME II.
GN   SCRA OR SPY1815.
OS   Streptococcus pyogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1314;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX   MEDLINE=21192684; PubMed=11296296;
RA   Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA   Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA   Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
RA   Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT   "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR   EMBL; AE006608; AAK34540.1; -.
DR   HSSP; P20166; 1GPR.
DR   InterPro; IPR000847; HTH_LysR.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
```

```
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.
SQ   SEQUENCE   620 AA;  65725 MW;  64F5FE83524DC4DC CRC64;


  Query Match             26.7%;  Score 632;  DB 16;  Length 620;
  Best Local Similarity   35.6%;  Pred. No. 1.3e-31;
  Matches  160;  Conservative   66;  Mismatches  202;  Indels   22;  Gaps    6;

Qy       1 MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLH 60
           : :|   | | : ||   : |::   : ||  |   ||||||||   |  : | :||:||
Db     189 LMLVSNELPNAWVVA---SGGDVKPLTFFGF-VPVVGYQGTVLPAFFVGLVGAKLEKWLH 244

Qy      61 KRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGL 120
           |::    | |:|| ||  :   |    |||    : :::  | | :      : ||: |
Db     245 KKVPEALDLLVTPFLTFAIMSTLGLFVIGPVFHSLENLVLAGTQAVLHLPFGIAGLIVGG 304

Qy     121 VYSPIVITGLHQSFPPIELELF-NQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGL 179
           :   ||:||:|  |  :| :|  | |     |  : |  ||  | |||    || |||||
Db     305 IQQLIVVTGIHHIFNFLEAQLIANTGKDPFNAYLTAATAAQAGATLAVAVKTKSTKLKGL 364

Qy     180 AGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGV 239
           |  | :||:|||||||||||||||    |  |:   |:|| :  || |      |      |
Db     365 AFPSTLSALLGITEPAIFGVNLRYPKVFVSGLIGGALGGWVAGLFGIAGTGFGITVLPGT 424

Qy     240 VSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEA 299
           :      :: :||  :|   :||  |  :|             |    |:||      |:
Db     425 LLYLNGQLLQYLVTMLVGLGVAFAIAYTWGY----------QDRETLPLPAVEVDQTADQ 474

Qy     300 PAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 359
           ||       : :|| |  : ||:||| :|:|| :| |:|| |    | ||| ||: : |
Db     475 PALAEE---TLYSPLNGTVVDLSAVSDPVFSSGAMGQGLAIKPEDNTLYSPVDGKVEIVF 531

Qy     360 PSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAI 419
           :|||:|: :        ::|:||| || :: |  |  |    |  || |:||  ||   |
Db     532 ETGHAYAITS----SQGAEVLLHIGIDTESMAGDGFESLVAVGQAVKKGDLLGHFDPSKI 587

Qy     420 KAAGYEVTTPIVVSNYKKTGPVNTYGLGEI 449
             || : || ::|||      |:    | :
Db     588 AEAGLDDTTMMIVSNIADYQSVDILAQGHV 617


RESULT  14
Q48408
ID   Q48408      PRELIMINARY;      PRT;   621 AA.
AC   Q48408;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CELLOBIOSE-SPECIFIC PTS PERMEASE.
GN   CASA.
OS   Klebsiella oxytoca.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=571;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=P2;
RX   MEDLINE=97176370; PubMed=9023916;
RA   Lai X., Davis F.C., Hespell R.B., Ingram L.O.;
RT   "Cloning of cellobiose phosphoenolpyruvate-dependent
RT   phosphotransferase genes: functional expression in recombinant
RT   Escherichia coli and identification of a putative binding region for
RT   disaccharides.";
RL   Appl. Environ. Microbiol. 63:355-363(1997).
DR   EMBL; U61727; AAB51563.1; -.
DR   HSSP; P20166; 1GPR.
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; 1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
SQ   SEQUENCE   621 AA;  65492 MW;  DF1DE2A2A7B81339 CRC64;


  Query Match             25.9%;  Score 612.5;  DB 2;  Length 621;
  Best Local Similarity   32.4%;  Pred. No. 2.2e-30;
  Matches  154;  Conservative   89;  Mismatches  181;  Indels   51;  Gaps   11;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLPVLVVSWILATIEKFLHK 61
           |:  | ::  :|  |:|| |    : | |: |    |  :|:|::: ||:   |||  :
Db     182 ALTHPLMIQAFD--ASMAPGAATEYFL-GIPVTFINYSSSVIPIILASWVSCWIEKRSNA 238

Qy      62 RLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLYDFGGPVGGLLFGLV 121
             |  :     || : | :   |||: |||   |:  :||:| | :|     : |   | :
Db     239 ILPSSMKNFFTPAICLAVVVPLTFLIIGPVATWLSQLLANGYQLIYQVAPWLAGAAMGAL 298

Qy     122 YSPIVITGLHQSFPPIEL-ELFNQGGSFIFATASMANIAQGAACLAVFFLAKSEKLKGLA 180
           :   || |||    |: :  |   |   :     | : |  | | :|   :  : | ||
Db     299 WQVCVIFGLHWGLIPLMINNLAVLGHDSMMPMLLPAVMGQVGAALGIFLRTRDARQKVLA 358

Qy     181 GASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALFNIKAVALGAAGFLGVV 240
           |::  : : |:|||||:|:|| || ||  |     ||||::  |:   |:  : ||  ::
Db     359 GSAVSAGIFGVTEPAIYGLNLPLRRPFIFGCVYWPIGGAMVG-FSDSHVSTYSFGFGNII 417

Qy     241 S---------IDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDATAAPVPAG 291
           :         |||               :  |||  :   ::          |    : ||
Db     418 TLAQMIPPEGIDAT--------------VWGGAAGMFASLII---------ACVLTLVAG 454

Qy     292 TTKAEAE------APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKG 345
             :: ||      |||   ||   | ||:||  :||  | |: |||| || ||||:|: |
Db     455 LPRSSAEQAAVVVAPASV-ND---ILAPMTGSVLALDQVPDSTFASGLLGQGVAIIPSVG 510

Qy     346 QLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGDEV 405
           ::::| ||::   | : ||  : :     | :::|:|:| ||| |:|  |    |:||::
Db     511 KVIAPFSGEVASIFQTKHAIGLLS----DSGIELLIHVGIDTVKLDGAPFTAHVKEGDKI 566

Qy     406 KAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAK 460
           |||:|| |||  ||  |||:: |||::||      ::      ::||  ||:|::
Db     567 KAGDLLLEFDRQAILDAGYDLATPIIISNSDDFRTLDMVSASAVDAGQPLLSVSR 621


RESULT  15
Q97J79
ID   Q97J79      PRELIMINARY;      PRT;   628 AA.
AC   Q97J79;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT.
GN   CAC1407.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007652; AAK79375.1; -.
```

```
DR   InterPro; IPR001127; PTS_EIIA.
DR   InterPro; IPR001996; PTS_EIIB.
DR   InterPro; IPR003352; PTS_EIIC.
DR   Pfam; PF00358; PTS_EIIA_1; 1.
DR   Pfam; PF00367; PTS_EIIB; 1.
DR   Pfam; PF02378; PTS_EIIC; 1.
DR   ProDom; PD001476; PTS_EIIB; 1.
DR   ProDom; PD002243; PTS_EIIA; 1.
DR   PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
DR   PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW   Complete proteome.
SQ   SEQUENCE   628 AA;  67460 MW;  7F715A1D2B9DE19A CRC64;


  Query Match             25.9%;  Score 612;  DB 16;  Length 628;
  Best Local Similarity   34.1%;  Pred. No. 2.4e-30;
  Matches  154;  Conservative   72;  Mismatches  179;  Indels   46;  Gaps   11;

Qy       2 AMVFPSLVNGYDVAATMAAGEMPMWSLF--------------GLDVAQAGYQGTVLPVLV 47
           ::|:|:|       :|:  |: |:::||              |: |    |  ||:|:::
Db     180 SLVYPTL-------STLMTGK-PLYTLFQGTIFASPVYVTFLGIPVILMSYSSTVIPIIL 231

Qy      48 VSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHGLQGLY 107
            |::   :||   | :       : |  |||:   |: | |||   | | :|  |   :|
Db     232 ASYVGVKVEKAFAKIIPDVVKTFLVPFCTLLVMVPLSLIVIGPISTWAGKLLGAGTLAIY 291

Qy     108 DFGGPVGGLLFGLVYSPIVITGLHQSFPPI---ELELFNQGGSFIFATASMANIAQGAAC 164
           :    : ||  |  :   || |||    ||    | : :     | |    |: ||
Db     292 NLSPILAGLFIGAFWQVFVIFGLHWGLVPIAMNNLSVLHYDP--ILAGTLGASFAQTGVV 349

Qy     165 LAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGGALIALF 224
           ||:    |: ||||:|  : :| : |:|||||:|| |  : || |    ||||| :
Db     350 LAILIKTKNVKLKGIALPAFISGIFGVTEPAIYGVTLPRKKPFIISCIGAAIGGGITGFM 409

Qy     225 NIKAVALGAAGFLGVVS-IDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRNGSIDPDA 283
             |   :|  |   : | | |  |       ||::  |:|      |  ::   |  | :
Db     410 GTKLWMMGGLGIFAIPSYIGAKGMDRGFYGAVMSVVISF----VVGFLIMFFAGFKDEE- 464

Qy     284 TAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPT 343
               |  ||| :     |     |:: :|| |:   || | |  |::| || |:|| |
Db     465 ----VKQETTKKK----NELVKQETLV-SPLKGKIKTLSEVKDEAFSTGSLGKGIAIEPE 515

Qy     344 KGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNGTHFNPLKKQGD 403
           :|:|||||  | :   ||:|||  : :        :||:|:| ||| | | :|  : ||||
Db     516 EGKLVSPVDGVLATLFPTGHAVGIIS----DKGTEILIHVGMDTVQLEGKYFKTILKQGD 571

Qy     404 EVKAGELLCEFDIDAIKAAGYEVTTPIVVSN 434
             ||||: : ||||  || ||| :|||:||:|
Db     572 HVKAGDTILEFDIPKIKKAGYTLTTPVVVTN 602


Search completed: September 25, 2002, 06:07:18
Job time: 504 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:02:04 ; Search time 1711.96 Seconds
(without alignments)
12038.745 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 ctcatggcatctgcgccgtt.....gttgaaaccttgagtgttcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	48.8	3.2	680	9	AL506262	AL506262 AL506262
c	2	45.8	3.0	473	10	T03017	T03017 FB20C8 Feta
	3	45	2.9	605	12	BH175021	BH175021 369 Liste
	4	41.8	2.7	681	12	CNS02EOD	AL193990 Tetraodon
	5	41.6	2.7	895	12	CNS0071A	AL066286 Drosophil
	6	41.4	2.7	477	10	BE997958	BE997958 EST429681
	7	41.4	2.7	559	9	AW574083	AW574083 EST316674
	8	41.4	2.7	601	10	BG580528	BG580528 EST482254
	9	41.4	2.7	627	10	BE998413	BE998413 EST430136
	10	40.8	2.7	884	12	CNS006U0	AL065923 Drosophil
c	11	40.8	2.7	983	12	CNS03CVD	AL238306 Tetraodon
	12	40.8	2.7	1101	12	CNS00LXJ	AL078875 Drosophil
c	13	40.2	2.6	970	12	CNS03H6V	AL243904 Tetraodon
	14	39.8	2.6	606	10	BE997957	BE997957 EST429680
c	15	39.2	2.6	532	10	BI856505	BI856505 603385596
c	16	39.2	2.6	993	12	CNS00L0P	AL067821 Drosophil
c	17	38.8	2.5	539	9	AI070146	AI070146 UI-R-Y0-1
	18	38.8	2.5	714	9	AW583970	AW583970 ia07h02.x
c	19	38.6	2.5	401	9	AI436816	AI436816 SWOv3MCAM
c	20	38.6	2.5	1204	10	BG391833	BG391833 602418142
c	21	38.4	2.5	560	9	BB643902	BB643902 BB643902
	22	38.4	2.5	1101	12	CNS00Z1D	AL097099 Drosophil
c	23	38.2	2.5	500	10	BF727921	BF727921 1000055A0
c	24	38.2	2.5	525	12	CNS025EN	AL181976 Tetraodon
	25	38.2	2.5	531	9	AW573719	AW573719 EST316310
	26	38.2	2.5	621	10	BE998387	BE998387 EST430110
c	27	38.2	2.5	652	9	AI982977	AI982977 wt46c03.x
	28	38.2	2.5	687	9	AW980716	AW980716 EST391869
	29	38.2	2.5	703	9	AW980735	AW980735 EST391888
	30	38.2	2.5	713	10	BG581619	BG581619 EST483354
	31	38.2	2.5	737	10	BG581892	BG581892 EST483628
	32	38.2	2.5	789	10	BG582423	BG582423 EST484167
	33	38.2	2.5	795	10	BG583630	BG583630 EST485382
c	34	37.8	2.5	654	9	AI812147	AI812147 605087H03
c	35	37.6	2.5	726	10	BM014664	BM014664 603640493
c	36	37.6	2.5	752	10	BM423398	BM423398 AGENCOURT
c	37	37.6	2.5	780	10	BI118076	BI118076 602867346
	38	37.6	2.5	1101	12	CNS00FXE	AL071370 Drosophil
c	39	37.4	2.4	460	9	AI318021	AI318021 SWOvAFCAP
	40	37.4	2.4	552	10	BE721200	BE721200 188296 MA
c	41	37.4	2.4	741	10	BI856144	BI856144 603382851
c	42	37.4	2.4	904	12	CNS03HA0	AL244017 Tetraodon
	43	37.2	2.4	421	9	AW173566	AW173566 xj08h03.x
	44	37.2	2.4	561	9	AI917973	AI917973 tz16d08.x
	45	37.2	2.4	594	9	AW440291	AW440291 hb89a03.x

ALIGNMENTS

RESULT 1
AL506262/c
LOCUS AL506262 680 bp mRNA linear EST 04-JAN-2001
DEFINITION AL506262 Hordeum vulgare Barke developing caryopsis (3.-15.DAP) Hordeum vulgare cDNA clone HY02I24T 5', mRNA sequence.
ACCESSION AL506262
VERSION AL506262.1 GI:12032477
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 680)
AUTHORS Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.
TITLE EST sequencing and analysis in barley
JOURNAL Unpublished (2000)
COMMENT Contact: Michalek W
Institute for Plant Genetics and Crop Plant Research
Corrensstr.3, D-06466 Gatersleben, Germany
Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
Seq primer: T3 primer for 5'end.
FEATURES Location/Qualifiers
source 1. .680
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY02I24T"
/clone_lib="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"
/tissue_type="developing caryopsis (3.-15.DAP)"
/lab_host="XLOLR"
/note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2: XhoI; mRNA was made from developing caryopsis (3.-15.DAP) of spring barley variety 'Barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb Sequence trimming: Vector sequences and sequence ends were

trimmed from the 5'-and 3'-end until a 50 bp window
contains less than two ambiguities. The maximum length was
set to 700 bp"
BASE COUNT 163 a 204 c 180 g 133 t
ORIGIN

Query Match 3.2%; Score 48.8; DB 9; Length 680;
Best Local Similarity 45.8%; Pred. No. 0.04;
Matches 250; Conservative 0; Mismatches 287; Indels 9; Gaps 2;

```
Qy    220 caccgtgcttcctgtgctggtggtttcttggattctggcaacgatcgagaagttcctgca 279
          ||| ||| | || | ||||||| |  | |||  ||| ||  |||||    |  || |
Db    551 CACGGTGATCCCGGCGCTGGTGATGACCTGGTGCCTGTCATATATCGAACGCTGGGTGGA 492

Qy    280 caagcgactcaagggcactgcagacttcctgatcactccagtgctgacgttgctgctcac 339
          ||          ||      |  |   | |  |||  ||| ||||||   || || |  |
Db    491 CAGCATTACACCGGCGGTGACCAAAAACTTCCTCAAACCAATGCTGATTGTGTTGATTGC 432

Qy    340 cggattccttacattcatcgccattggcccagcaatgcgctgggtgggcgatgtgctggc 399
          || |   ||  || || |    ||||||||         |||| | ||    |   |  |
Db    431 CGCACCGCTGGCAATCCTGCTGATTGGCCCGATTGGTATCTGGATCGGTAGCGCCATTTC 372

Qy    400 acacggtctacagggactttatgatttcggtggtccagtcggcggtctgctcttcggtct 459
              |  |     || ||| ||     ||      |     |      |  | ||
Db    371 GGCGCTGGTTTACACCATTCATGGTTATCTGGGCTGGCTTTCAGTCGCCATTATGGGCGC 312

Qy    460 ggtctactcaccaatcgtcatcactggtctgcaccagtccttcccgccaattgagctgga 519
          | | |   | |   | || || || ||  ||||||    |||  ||||||      |  |
Db    311 GCTGTGGCCTCTGCTGGTAATGACCGGGATGCACCGCGTCTTTACGCCAACCATCATTCA 252

Qy    520 gctgtttaaccagggtggatccttcatcttcgcaacggcatctatg------gctaatat 573
          |    ||  | |    ||          | |  | | | ||   |      |||||  |
Db    251 GACCATTGCCGAAACCGGCAAAGAAGGGATGGTCATGCCGTCAGAGATCGGCGCTAACCT 192

Qy    574 cgcccagggtgcggcatgtttggcagtgttcttcctggcgaagagtgaaaagctcaaggg 633
            | | ||| |   |||   ||||||||   |      |||| |      | ||
Db    191 GTCGCTGGGCGGTTCATCACTGGCAGTGGCGTGGAAAACGAAAAACCCGGAACTGCGTCA 132

Qy    634 ccttgcaggtgcttcaggtgtctccgctgttct---tggtattacggagcctgcgatctt 690
            ||    ||| | |   |  || ||  || |    || ||| | || || ||| | |
Db    131 GACGGCGCTGGCTGCGGCGGCATCAGCCATTATGGCGGGGATTTCCGAACCGGCGTTATA 72

Qy    691 cggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagctatcgg 750
          || |||    | ||||||    | ||| |  ||  | || |    | ||    |||  |
Db     71 TGGCGTGGCGATCCGCCTGAAACGTCCGCTTATCGCCAGTCTTATCAGCGGTTTTATTTG 12

Qy    751 tggcgc 756
          ||||||
Db     11 CGGCGC 6
```

RESULT 2
T03017/c
LOCUS T03017 473 bp mRNA linear EST 14-FEB-1997
DEFINITION FB20C8 Fetal brain, Stratagene Homo sapiens cDNA clone FB20C8, mRNA sequence.
ACCESSION T03017
VERSION T03017.1 GI:314258
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS Khan,A.S., Wilcox,A.S., Polymeropoulos,M.H., Hopkins,J.A., Stevens,T.J., Robinson,M., Orpana,A.K. and Sikela,J.M.
TITLE Single pass sequencing and physical and genetic mapping of human cDNAs
JOURNAL Nature Genet. 2, 180-185 (1992)
MEDLINE 94258200
COMMENT On Sep 21, 1992 this sequence version replaced gi:279156.
Contact: Sikela JM
Department of Pharmacology
University of Colorado Health Sciences Center
Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
Tel: 3032708637
Fax: 3032707097
Email: nikki@tally.uchsc.edu.
FEATURES Location/Qualifiers
source 1. .473
/organism="Homo sapiens"
/db_xref="ATCC (inhost):86662"
/db_xref="GDB:D0S2785E"
/db_xref="taxon:9606"
/clone="FB20C8"
/clone_lib="Fetal brain, Stratagene"
/note="Vector: Lambda ZAP II; The FB library (catalog #937201, Stratagene) was constructed by directional cloning and oligo d(T)-priming in Lambda ZAP II phage utilizing mRNA from a 17 week human fetus (total brain)."
BASE COUNT 121 a 124 c 104 g 123 t 1 others
ORIGIN

Query Match 3.0%; Score 45.8; DB 10; Length 473;
Best Local Similarity 59.7%; Pred. No. 0.21;
Matches 77; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy   1231 ggatatcttgatgcacattggtttcgacacagtaaacctcaacggcacgcactttaaccc 1290
          ||| ||| |  | || || ||| ||||||| |||   ||  | || |    ||||||  |
Db    275 GGAGATCGTCGTCCATATGGGTATCGACACCGTAGCGCTGGAAGGTAAAGGCTTTAAACG 216

Qy   1291 gctgaagaagcagggcgatgaagtcaaagcaggggagctgctgtgtgaattcgatattga 1350
           |||  | |  |||| |    | || |  ||||||  | | | |    ||| | ||| | ||
Db    215 TCTGGTGGAAGAGGGTGCGCAGGTAAGCGCAGGGCAACCGATTCTGGAAATGGATCTGGA 156

Qy   1351 tgccattaa 1359
          |  | | ||
Db    155 TTACCTGAA 147
```

RESULT 3
BH175021
LOCUS BH175021 605 bp DNA linear GSS 01-JAN-2002
DEFINITION 369 Listeria monocytogenes pUC18 library Listeria monocytogenes genomic clone 63-43 similar to PTS system beta-glucoside-specific enzymeII, DNA sequence.
ACCESSION BH175021
VERSION BH175021.1 GI:18030243
KEYWORDS GSS.
SOURCE Listeria monocytogenes.
ORGANISM Listeria monocytogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales; Listeriaceae; Listeria.
REFERENCE 1 (bases 1 to 605)
AUTHORS Borucki,M.K., Krug,M.J., Muraoka,W.T. and Call,D.R.
TITLE Genetic characterization of Listeria monocytogenes using DNA microarrays
JOURNAL Unpublished (2001)
COMMENT Contact: Monica Borucki
Animal Disease Research Unit
USDA-ARS
3003 ADBF, WSU, Pullman, WA 99164, USA
Tel: 509 335 7407
Fax: 509 335 8328
Email: mborucki@vetmed.wsu.edu
PCR PRimers
FORWARD: M13F
BACKWARD: M13R
Insert Length: 605 Std Error: 0.00
Seq primer: M13F, M13R
Class: plasmid ends.
FEATURES Location/Qualifiers
source 1. .605
/organism="Listeria monocytogenes"

```
                     /strain="multiple"
                     /db_xref="taxon:1639"
                     /clone="63-43"
                     /clone_lib="Listeria monocytogenes pUC18 library"
                     /note="Vector: pUC18; Shotgun library prepared by Amplicon
                     Express (Pullman, WA)"
BASE COUNT       154 a    110 c    125 g    216 t
ORIGIN


  Query Match             2.9%;  Score 45;  DB 12;  Length 605;
  Best Local Similarity   46.8%;  Pred. No. 0.39;
  Matches  141;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy     210 gttaccagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgatcgaga 269
           ||||      | ||||  ||||  |  |  |||    |||| || |  | |   | || |
Db     181 GTTATTCTTTCTCCGTTATTCCAATTATCTTGGCAGTTTGGTTTTTATCTATTTTAGAAA 240

Qy     270 agttcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgctgacgt 329
             || ||  | |    | |  | |   | ||  |   | |  | || ||  || |
Db     241 GATTTCTTAATAGTAAAATTCATGAGGCAGCGAAGACCTTTTTAACGCCGATGATTTGCC 300

Qy     330 tgctgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggtgggcg 389
           || |||| | || |   |||||||||| | ||  |||| ||          |  | |
Db     301 TGATGCTTATCGTACCGCTTACATTCTTAGCGTTTGGACCTCTGGGTACGTTTATTAGTC 360

Qy     390 atgtgctggcacacggtctacagggactttatgatttcggtggtccagtcggcggtctgc 449
           | |   | ||   |||           ||||| | ||  |        || |  |||
Db     361 AAGGTTTAGCTAGCGGGTATACATTTATTTATAACTTGAGCCCAATTGTAGCGGGTGCTT 420

Qy     450 tcttcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttcccgccaa 509
           |  | |||  | |||    |  | | || ||   |||  | ||   |  |||   |||||
Db     421 TTATGGGTGCGTTCTGGCAAGTATTAGTTATTTTTGGGATTCATTGGGGCTTTGTGCCAA 480

Qy     510 t 510
           |
Db     481 T 481


RESULT   4
CNS02EOD
LOCUS       CNS02EOD                681 bp    DNA     linear   GSS 13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            262H14 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL193990
VERSION     AL193990.1  GI:7832096
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 681)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 681)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="262H14"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG262DD07LP1-end : T7"
BASE COUNT       74 a    105 c    170 g    282 t     50 others
ORIGIN


  Query Match             2.7%;  Score 41.8;  DB 12;  Length 681;
  Best Local Similarity   40.9%;  Pred. No. 3;
  Matches  139;  Conservative   18;  Mismatches  183;  Indels    0;  Gaps    0;

Qy     632 ggccttgcaggtgcttcaggtgtctccgctgttcttggtattacggagcctgcgatcttc 691
           | | |||  | || ||  | |||    ||:| : |:| | |:   |    ||   | |
Db      58 GCCGTTGTYGTTGTTTTTGTTGTTGTTGCYGCYGTYGTTGTYGTTGTTTTTGTTGTTGTT 117

Qy     692 ggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagctatcggt 751
           | |||     ||    |  :       |  ||  :| : | | |  :|    | |:| |
Db     118 GTTGTTGTTTTTGTTGTTGTYGTTGTTGTTGTCGCYGTYGTTGTTGTYGTTTTTGTYGTT 177

Qy     752 ggcgctttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtttcttg 811
           |  | | |  | ||  |  ||  |  :   |  |||    | |  | ||  | | |  |
Db     178 GTNGTTGTTGTNGCCGTNGTTGTTGCYGTTGTTGTTTTTGTTGTTGTTGTTGTTGTY 237

Qy     812 ggtgttgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagttgttacc 871
           |  |  ||   : ||| :| |  | | | ||| |  || | ||   ||| ||:|||  |
Db     238 GTCGCCGTCGTYGTTGCYGTTTTTGTTGTCGTCGTCGTCGTCGCTGCCGTYGTTGTC 297

Qy     872 ttcttcatcgcattcggcgcagcgattgcttatggcctttacttggttcgccgcaacggc 931
             || |||    || :|  |   ||| |   | | |:  | | |||     |  || |
Db     298 GCTGTCGTCGTCGCCGTYGTCGTTTTTGTTGTCGTCGTYGTCGTCGTTGTTGTCGCCGCC 357

Qy     932 agcattgatccagatgcaaccgctgctccagtgcctgcag 971
           |||  :   | || |   | :| |   ||   || :|
Db     358 GTTGTTGTWTTTGTTGTAGTWGTYGTTTTTGTTGTTGTWG 397


RESULT   5
CNS0071A
LOCUS       CNS0071A                895 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066286
VERSION     AL066286.1  GI:4945153
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 895)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
```

```
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .895
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR14B09"
                      /note="end : TET3"
BASE COUNT      124 a     80 c    204 g    179 t    308 others
ORIGIN


  Query Match             2.7%;  Score 41.6;  DB 12;  Length 895;
  Best Local Similarity   19.9%;  Pred. No. 4;
  Matches   68;  Conservative  115;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     580 gggtgcggcatgtttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgc 639
           |:|:: ::   :::|::  :|::|  :  ||:    : : :    | :   | |  ||:
Db     512 GKGKKGKKKGGKKKTKKKKKTKKTGGKKGTGKGTKKKKKGKKTKTGGKKKKGTGTKTTKK 571

Qy     640 aggtgcttcaggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaa 699
            : :| |:    |:|: :  : ||:| ::|   |:  :  :  :: | | ::  | |:|
Db     572 KKTKGKTKKKTGKGKTKKKKGTGKTGKKGTGKTKKTKTGKGKKKGGGTGKKGTGGGKGGK 631

Qy     700 ccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgcttt 759
             :| :  || :  |:  :|: :: : ::: :  ||   :  : :   :::::: : |:|
Db     632 KGKTKKTGTGKKTGGKTKKTKKKKKTKKKKKKGTGTKGGKKGKGKTGKKKKKKKKGTKT 691

Qy     760 gattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgt 819
           : ||    | |:|  | :   ||  | :| ||:|:: | :|  |:  : :: : ::|:
Db     692 KKTTTKTTTTTKTKTTGKKKTGGKGGGKGKGTKGKKKGTKGGGGKGGKKKKTKTKKTKTG 751

Qy     820 ttctattgatgctccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcat 879
           :  : |:  :: |   : :  |:::     : :|:::|||::  : :|:    :: :|  :
Db     752 KGKKKTKTTKKKTGKTKKKGTKKKKTGTGKGKTKKKGTGKKGTKGKGKGGKKKKKKTGKK 811

Qy     880 cgcattcggcgcagcgattgcttatggccctttacttggttcg 921
             |: |: |  :   :::  : | :  : |  ::::|: |
Db     812 TTTGTKGGKKGGGKKTKKKKKGKKTTKKKKGTKTKKKKTKKG 853


RESULT   6
BE997958
LOCUS       BE997958                 477 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  EST429681 GVSN Medicago truncatula cDNA clone pGVSN-8B9, mRNA
            sequence.
ACCESSION   BE997958
VERSION     BE997958.1  GI:10698234
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
            ,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
  TITLE     ESTs from senescent nodules of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M271936e TIGR sequence name:
            MTKAK05TKB More information is available at:
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .477
                      /organism="Medicago truncatula"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="pGVSN-8B9"
                      /clone_lib="GVSN"
                      /tissue_type="senescent root nodules"
                      /dev_stage="mixture of effective nodules from 40 day old
                      plants harvested 36 hours post shoot removal and nodules
                      collected from 2 month old plants at mid-pod stage"
                      /lab_host="E. coli strain SOLR"
                      /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from the
                      mixture of effective nodules of 40 day old plants
                      harvested 36 hours post shoot removal and nodules
                      collected from 2 month old plants at mid-pod stage. The
                      cDNA was directionally ligated into the Uni-ZAP XR vector
                      from Stratagene and packaged using Gigapack III Gold
                      packaging extracts. Plasmids containing cDNA inserts were
                      excised from the recombinant lambda-ZAP phage using
                      Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT      137 a    100 c    101 g    139 t
ORIGIN


  Query Match             2.7%;  Score 41.4;  DB 10;  Length 477;
  Best Local Similarity   50.2%;  Pred. No. 3.3;
  Matches  102;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    1246 cattggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcaggg 1305
           |||||  |||  || || | | ||   |   |   || ||| ||    | | | ||
Db     258 CATTGACTTCACCAAAGAAGAGCTTCTCATAGCTTTCTCTAAGCCTAACACGGATCATAA 317

Qy    1306 cgatgaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgc 1365
           | |   ||||  |||   ||| ||| | | | ||    | | | |||||||||  |
Db     318 TGGTTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAGGCGAC 377

Qy    1366 aggttatgaggtaaccacgccgattgttgtttcgaattacaagaaaaccggacctgtaaa 1425
             | | || || | |        ||  ||   |  |||   | |     |||| | |||
Db     378 CAATGAAGAAGTGAGCGACTTTGTTCGTGAGGCTGATTCTGATAGTGATGGACATCTAAG 437

Qy    1426 cacttacggtttgggcgaaattg 1448
           |  | | |  || | |  | |||
Db     438 CTTTAAGGAGTTTGTCAGACTTG 460


RESULT   7
AW574083
LOCUS       AW574083                 559 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  EST316674 GVN Medicago truncatula cDNA clone pGVN-51G2, mRNA
            sequence.
ACCESSION   AW574083
VERSION     AW574083.1  GI:7238816
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng
            ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from one month old nitrogen-fixing root nodules of Medicago
            truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
```

```
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            Minnesota EST name:M252981e
            TIGR sequence name:MTCAR37TK
            More information is available at. .
            'http://chrysie.tamu.edu/medicago'
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .559
                      /organism="Medicago truncatula"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="pGVN-51G2"
                      /clone_lib="GVN"
                      /tissue_type="N2-fixing root nodules"
                      /dev_stage="effective root nodules harvested one month
                      post inoculation with Sinorhizobium meliloti"
                      /lab_host="E. coli strain XLOLR"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from
                      effective root nodules harvested one month post
                      inoculation with Sinorhizobium meliloti. The cDNA was
                      directionally ligated into the Uni-ZAP XR vector from
                      Stratagene and packaged using Gigapack III Gold packaging
                      extracts. Plasmids containing cDNA inserts were excised
                      from the recombinant lambda-ZAP phage using Ex-Assist
                      helper phage and propagated in XLOLR cells."
BASE COUNT      159 a    112 c    112 g    176 t
ORIGIN


  Query Match             2.7%;  Score 41.4;  DB 9;  Length 559;
  Best Local Similarity   50.2%;  Pred. No. 3.5;
  Matches  102;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy   1246 cattggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcaggg 1305
           ||||| ||| || || | | ||   |    |   || ||| ||    | | | ||
Db    316 CATTGACTTCACCAAAGAAGAGCTTCTCATAGCTTTCTCTAAGCCTAACACGGATCATAA 375

Qy   1306 cgatgaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgc 1365
           | |   |||| |||   ||| ||| | | | ||    | | | | |||||||||  |
Db    376 TGGTTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAGGCGAC 435

Qy   1366 aggttatgaggtaaccacgccgattgttgtttcgaattacaagaaaaccggacctgtaaa 1425
             | | || || | |       ||  ||   |  |||   | |     |||| | |||
Db    436 CAATGAAGAAGTGAGCGACTTTGTTCGTGAGGCTGATTCTGATAGTGATGGACATCTAAG 495

Qy   1426 cacttacggtttgggcgaaattg 1448
           |  | | |  || | |  | |||
Db    496 CTTTAAGGAGTTTGTCAGACTTG 518


RESULT   8
BG580528
LOCUS       BG580528                601 bp    mRNA    linear   EST 11-APR-2001
DEFINITION  EST482254 GVN Medicago truncatula cDNA clone pGVN-57H8 5' end, mRNA
            sequence.
ACCESSION   BG580528
VERSION     BG580528.1  GI:13595592
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town
            ,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
  TITLE     ESTs from one month old nitrogen-fixing root nodules of Medicago
            truncatula, 2001
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M381580e TIGR sequence name:
            MTCCT40TK More information is available at: http://www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .601
                      /organism="Medicago truncatula"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="pGVN-57H8"
                      /clone_lib="GVN"
                      /tissue_type="N2-fixing root nodules"
                      /dev_stage="effective root nodules harvested one month
                      post inoculation with Sinorhizobium meliloti"
                      /lab_host="E. coli strain XLOLR"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from
                      effective root nodules harvested one month post
                      inoculation with Sinorhizobium meliloti. The cDNA was
                      directionally ligated into the Uni-ZAP XR vector from
                      Stratagene and packaged using Gigapack III Gold packaging
                      extracts. Plasmids containing cDNA inserts were excised
                      from the recombinant lambda-ZAP phage using Ex-Assist
                      helper phage and propagated in XLOLR cells."
BASE COUNT      170 a    121 c    127 g    183 t
ORIGIN


  Query Match             2.7%;  Score 41.4;  DB 10;  Length 601;
  Best Local Similarity   50.2%;  Pred. No. 3.7;
  Matches  102;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy   1246 cattggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcaggg 1305
           ||||| ||| || || | | ||   |    |   || ||| ||    | | | ||
Db    307 CATTGACTTCACCAAAGAAGAGCTTCTCATAGCTTTCTCTAAGCCTAACACGGATCATAA 366

Qy   1306 cgatgaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgc 1365
           | |   |||| |||   ||| ||| | | | ||    | | | | |||||||||  |
Db    367 TGGTTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAGGCGAC 426

Qy   1366 aggttatgaggtaaccacgccgattgttgtttcgaattacaagaaaaccggacctgtaaa 1425
             | | || || | |       ||  ||   |  |||   | |     |||| | |||
Db    427 CAATGAAGAAGTGAGCGACTTTGTTCGTGAGGCTGATTCTGATAGTGATGGACATCTAAG 486

Qy   1426 cacttacggtttgggcgaaattg 1448
           |  | | |  || | |  | |||
Db    487 CTTTAAGGAGTTTGTCAGACTTG 509


RESULT   9
BE998413
LOCUS       BE998413                627 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  EST430136 GVSN Medicago truncatula cDNA clone pGVSN-9J12, mRNA
            sequence.
ACCESSION   BE998413
VERSION     BE998413.1  GI:10698689
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
            ,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
```

```
  TITLE     ESTs from senescent nodules of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M272391e TIGR sequence name:
            MTKAP54TK More information is available at:
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pGVSN-9J12"
                     /clone_lib="GVSN"
                     /tissue_type="senescent root nodules"
                     /dev_stage="mixture of effective nodules from 40 day old
                     plants harvested 36 hours post shoot removal and nodules
                     collected from 2 month old plants at mid-pod stage"
                     /lab_host="E. coli strain SOLR"
                     /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from the
                     mixture of effective nodules of 40 day old plants
                     harvested 36 hours post shoot removal and nodules
                     collected from 2 month old plants at mid-pod stage. The
                     cDNA was directionally ligated into the Uni-ZAP XR vector
                     from Stratagene and packaged using Gigapack III Gold
                     packaging extracts. Plasmids containing cDNA inserts were
                     excised from the recombinant lambda-ZAP phage using
                     Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT      175 a    124 c    135 g    193 t
ORIGIN
```

```
  Query Match             2.7%;  Score 41.4;  DB 10;  Length 627;
  Best Local Similarity   50.2%;  Pred. No. 3.7;
  Matches  102;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy   1246 cattggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcaggg 1305
          ||||| ||| || || | | ||   |     |   || ||| ||    | | | ||
Db    313 CATTGACTTCACCAAAGAAGAGCTTCTCATAGCTTTCTCTAAGCCTAACACGGATCATAA 372

Qy   1306 cgatgaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgc 1365
           | |   ||||  |||   ||| |||  | | | ||    | | | | |||||||||  |
Db    373 TGGTTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAGGCGAC 432

Qy   1366 aggttatgaggtaaccacgccgattgttgtttcgaattacaagaaaaccggacctgtaaa 1425
            | | || || | |       ||  ||   |  |||   | |     |||| | |||
Db    433 CAATGAAGAAGTGAGCGACTTTGTTCGTGAGGCTGATTCTGATAGTGATGGACATCTAAG 492

Qy   1426 cacttacggtttgggcgaaattg 1448
          |  | | |  || | |  | |||
Db    493 CTTTAAGGAGTTTGTCAGACTTG 515
```

```
RESULT  10
CNS006U0
LOCUS       CNS006U0                  884 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065923
VERSION     AL065923.1  GI:4944891
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14N21"
                     /note="end : T7"
BASE COUNT      230 a     62 c    139 g    124 t    329 others
ORIGIN
```

```
  Query Match             2.7%;  Score 40.8;  DB 12;  Length 884;
  Best Local Similarity   15.0%;  Pred. No. 6.5;
  Matches   55;  Conservative  147;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    572 atcgcccagggtgcggcatgtttggcagtgttcttcctggcgaagagtgaaaagctcaag 631
          :: :: :: ::|| ||  ::| |::  :  :   :    | :  :  |:::  | :   |
Db    516 RKGKVGMVAVKTGGGGGGKRTGTKKGGKSTKGDGKKGVTGGKTKKTTTKDDTTGKKTTGG 575

Qy    632 ggccttgcaggtgcttcaggtgtctccgctgttcttggtattacggagcctgcgatcttc 691
            |  :||  ||:| |    | |: :  : | :| ||: : |:   :    ||   | ::
Db    576 TGAGKTGGKGGKGGTGGGTGGGKGKGGKTTTKTKTTKTKGTKTKTKTTGKTGTTGTKKKK 635

Qy    692 ggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagctatcggt 751
          || | |    :: :  || |  ||  ::  |:    :||   ::|   :  : : | |:
Db    636 GGGGGGGGKKKKGKGGTGKGGGGGGKKKGGTKTGGKDGTGGKKKTTKKKKKKKKGTKGKG 695

Qy    752 ggcgctttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtttcttg 811
          :|   :: | :::   | : |  : |   ::  |: : |:::|| | :|  ::: : :::
Db    696 KGKTKKKSGTKKKKGGTGKGTGTKGTGTKKKKGGKGKGGKKKGGKGKKGKKKKKGKKKKK 755

Qy    812 ggtgttgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagttgttacc 871
          :::: :::::   :|: ::     : : :|:: :::: :::::::::  ::::::
Db    756 KKKKGKKKKKTGKKTKGKKGGTKKKGKKKGKKTKKKKKKKKKKKKKKKKKKKKKKKKKK 815

Qy    872 ttcttcatcgcattcggcgcagcgattgcttatggcctttacttggttcgccgcaacggc 931
          :: ::  : :  :: :: :  : : ::: :: :::  :::  :: ::: :   :::: :
Db    816 KKKKKKKKTKKKKKKKKKKCKKKKKKKKKKKKKKKKKKKKKKNKBKKDAAMMMMMADK 875

Qy    932 agcattg 938
          :: :::
Db    876 KKBKKKK 882
```

```
RESULT  11
CNS03CVD/c
LOCUS       CNS03CVD                  983 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            015D07 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL238306
```

```
VERSION     AL238306.1  GI:7897441
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 983)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 983)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="015D07"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG015CB04SP1~end :
                     PUC-Ori"
BASE COUNT      202 a    289 c    274 g    201 t     17 others
ORIGIN
```

```
  Query Match             2.7%;  Score 40.8;  DB 12;  Length 983;
  Best Local Similarity   53.0%;  Pred. No. 6.8;
  Matches   87;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy    705 gcctgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattg 764
          ||| | || ||  | |   |  | |  |||  |||  ||| | || |     ||
Db    732 GCCGGAGCCGGGAGCTGGCCCCCCGCTCCGGCGCCGGCGCTGTTGGGGTTTTATTACCGT 673

Qy    765 cactctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttcta 824
          | |||| |||||     ||||   | |  | ||  | | |  | | |||||||  |
Db    672 TGTTATTTATTATCATTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 613

Qy    825 ttgatgctccagatatggtcatgttcttggtgtgtgcagttgtt 868
          |||  |||    | | | ||  || |  |||||  ||  ||||||
Db    612 TTGTTGCGGTTGTTGTTGTTGTGGTGGTGGTGGTTGTTGTTGTT 569
```

```
RESULT  12
CNS00LXJ
LOCUS       CNS00LXJ                1101 bp    DNA     linear   GSS 14-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR48E16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078875
VERSION     AL078875.1  GI:5102165
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR48E16"
                     /note="end : T7"
BASE COUNT      158 a    208 c    167 g    310 t    258 others
ORIGIN
```

```
  Query Match             2.7%;  Score 40.8;  DB 12;  Length 1101;
  Best Local Similarity   24.5%;  Pred. No. 7.2;
  Matches   79;  Conservative   96;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    292 gggcactgcagacttcctgatcactccagtgctgacgttgctgctcaccggattccttac 351
          |||  |  |   | | |    : | |: |  :|  |  |: ||::| :    |:  |  :
Db    629 GGGKTCCTCCCCCCTGCCTTCSCCCCSSGCYYTTYCCCTKTTGSKCTBTYTGTYTTTSCS 688

Qy    352 attcatcgccattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctaca 411
            |: ||:|:  :::::: ::  :| |: :|::| :|: | |:  ||    :|:||: :
Db    689 CGTSTTCSCSSSBSSSSSSSSSGKGGGSGKGBKGTSCBTTTTSTCGGGGTGSGKTCBCBK 748

Qy    412 gggactttatgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcacc 471
            |  :| | :    | | : ::|:    ||  ::     | |::  :|: | | :|: |:
Db    749 TGTTYTSTGYYKCGTKGTKCBBCSBTGTGGGTSKTGYKCCSTBSCCSTSTTKTGSTBKCB 808

Qy    472 aatcgtcatcactggtctgcaccagtccttcccgccaattgagctggagctgtttaacca 531
            :: :: :: |       : || : :|:::|| ::    :  | |:  :| :|   ::
Db    809 GCKSYYYGBYCCSTTGTCYYSCCCKCYCYYYCCYBYBCCSSCCCSGSSCSTYKTSCSSST 868

Qy    532 gggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatg 591
          | |||:  :  :  |   :::  :| |  |:  : :     |:::::  |:||::| : :
Db    869 GCGTGSGSSTCBKYTKSGYSYYTSGCCSSCYCCBTSSGGSKTSSSSSCGBGTKYGYYCBY 928

Qy    592 tttggcagtgttcttcctggcg 613
          ::| :| : ::|| :|  :: |
Db    929 YYTCBCCKGKYTCCYCTCKBGG 950
```

```
RESULT  13
CNS03H6V/c
LOCUS       CNS03H6V                970 bp    DNA     linear   GSS 17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            026011 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL243904
VERSION     AL243904.1  GI:7964916
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 970)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 970)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 970)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .970
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="026011"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG026AH06LP1~end : T7"
BASE COUNT      311 a    298 c    194 g    152 t     15 others
ORIGIN
```

```
  Query Match             2.6%;  Score 40.2;  DB 12;  Length 970;
  Best Local Similarity   45.9%;  Pred. No. 9.8;
  Matches  130;  Conservative    2;  Mismatches  151;  Indels    0;  Gaps    0;

Qy     637 tgcaggtgcttcaggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgt 696
           ||| | || |   | ||   | | |||| ||| | |  | |   |||   |  | | |||
Db     709 TGCTGCTGTTGTTGCTGCTGCTGTTGTTGTTGTTGTCGCTGCTGCTGTTGTTGTTGCTGT 650

Qy     697 gaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgc 756
           |              |  |:   | | | | |   | ||||   | ||  |
Db     649 TGTTGCTGCTGCTGTTGTTGTTGTTGTVGCTGCTGTTGTTGATGTAGCTGCTGTTGTTGT 590

Qy     757 tttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgt 816
           |    |||  |  ||  | |:   || ||||   | |  ||||| | | |  | | ||
Db     589 TGCTGCTGCTGTTGTTGTTGTMGCTGCTGTTGTTGTTGCTGCTGCTGTTGTTGTTGCTGC 530

Qy     817 tgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagttgttaccttctt 876
           || |  | ||| |||| | | | | ||        | ||   ||| ||||||  |  |  |
Db     529 TGCTGTTGTTGTTGCTGCTGCTGTTGTTGCTGCTGTTGTTGCTGCTGTTGTTGCTGTTGT 470

Qy     877 catcgcattcggcgcagcgattgcttatggcctttacttggtt 919
             | ||    |  || |  |||||  ||   ||    | |||
Db     469 TGTTGCTGCTGCTGCTGTTGTTGCTGCTGTTGTTGCTGTTGTT 427
```

```
RESULT  14
BE997957
LOCUS       BE997957                606 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  EST429680 GVSN Medicago truncatula cDNA clone pGVSN-8B9, mRNA
            sequence.
ACCESSION   BE997957
VERSION     BE997957.1  GI:10698233
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
            ,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
  TITLE     ESTs from senescent nodules of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M271935e TIGR sequence name:
            MTKAK05TK More information is available at:
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pGVSN-8B9"
                     /clone_lib="GVSN"
                     /tissue_type="senescent root nodules"
                     /dev_stage="mixture of effective nodules from 40 day old
                     plants harvested 36 hours post shoot removal and nodules
                     collected from 2 month old plants at mid-pod stage"
                     /lab_host="E. coli strain SOLR"
                     /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from the
                     mixture of effective nodules of 40 day old plants
                     harvested 36 hours post shoot removal and nodules
                     collected from 2 month old plants at mid-pod stage. The
                     cDNA was directionally ligated into the Uni-ZAP XR vector
                     from Stratagene and packaged using Gigapack III Gold
                     packaging extracts. Plasmids containing cDNA inserts were
                     excised from the recombinant lambda-ZAP phage using
                     Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT      173 a    122 c    127 g    184 t
ORIGIN
```

```
  Query Match             2.6%;  Score 39.8;  DB 10;  Length 606;
  Best Local Similarity   49.8%;  Pred. No. 9.9;
  Matches  101;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy    1246 cattggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcaggg 1305
           |||||  |||  || || | | ||   |     |   || ||| ||    | | | ||
Db     317 CATTGACTTCACCAAAGAAGAGCTTCTCATAGCTTTCTCTAAGCCTAACACGGATCATAA 376

Qy    1306 cgatgaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgc 1365
           | |   ||||  |||   ||| ||| | | | ||    | | | | |||||| ||  |
Db     377 TGGTTTTGTCACTGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAAGCGAC 436

Qy    1366 aggttatgaggtaaccacgccgattgttgtttcgaattacaagaaaaccggacctgtaaa 1425
           | | || || | |      ||  ||   |  |||   | |     |||| | |||
Db     437 CAATGAAGAAGTGAGCGACTTTGTTCGTGAGGCTGATTCTGATAGTGATGGACATCTAAG 496

Qy    1426 cacttacggtttgggcgaaattg 1448
           |  | | |  || | |  | |||
Db     497 CTTTAAAGAGTTTGTCAGACTTG 519
```

```
RESULT  15
BI856505/c
LOCUS       BI856505                532 bp    mRNA    linear   EST 10-OCT-2001
DEFINITION  603385596F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5394670 5',
            mRNA sequence.
ACCESSION   BI856505
VERSION     BI856505.1  GI:15997252
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12005  row: o  column: 23
            High quality sequence stop: 522.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5394670"
                     /clone_lib="NIH_MGC_87"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      185 a    122 c    134 g     91 t
ORIGIN


  Query Match             2.6%;  Score 39.2;  DB 10;  Length 532;
  Best Local Similarity   47.2%;  Pred. No. 13;
  Matches  119;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy     757 tttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgt 816
           | |||||||    | |  | ||  | ||| ||| |  || || || | ||||| || ||
Db     472 TATGATTGCCACTTATGCTTTCTTGCCAGATGCCTCTGGAGCAGCTGCTTTCTGGGCTGG 413

Qy     817 tgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagttgttaccttctt 876
           | ||||          || |   |        |   | |       |||  | ||||
Db     412 AGCTTCTTGACCCTTCTGAGCTTCTGGAGCAGGCGCTGCTTTCTGGCCTGTGGCTTTCTG 353

Qy     877 catcgcattcggcgcagcgattgcttatggcctttacttggttcgccgcaacggcagcat 936
               | |  |  |   ||    || | |   ||      |||   |       |   | |
Db     352 GGCAGGAACCTTCTGGGCTGGAGCCTTTTTACTCGCGGCGGTGATCTTTTTTGCTGGAAC 293

Qy     937 tgatccagatgcaaccgctgctccagtgcctgcaggaacgaccaaagccgaagcagaagc 996
           |    |||  ||| | || ||  |||   | |||| | |  |   ||| | ||||| |||
Db     292 TTTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 233

Qy     997 acccgcagaatt 1008
           | | |||| | |
Db     232 AGCAGCAGCAGT 221


Search completed: September 25, 2002, 06:46:50
Job time: 2686 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:07:24 ; Search time 54.45 Seconds
(without alignments)
6888.565 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 ctcatggcatctgcgccgtt.....gttgaaaccttgagtgttcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

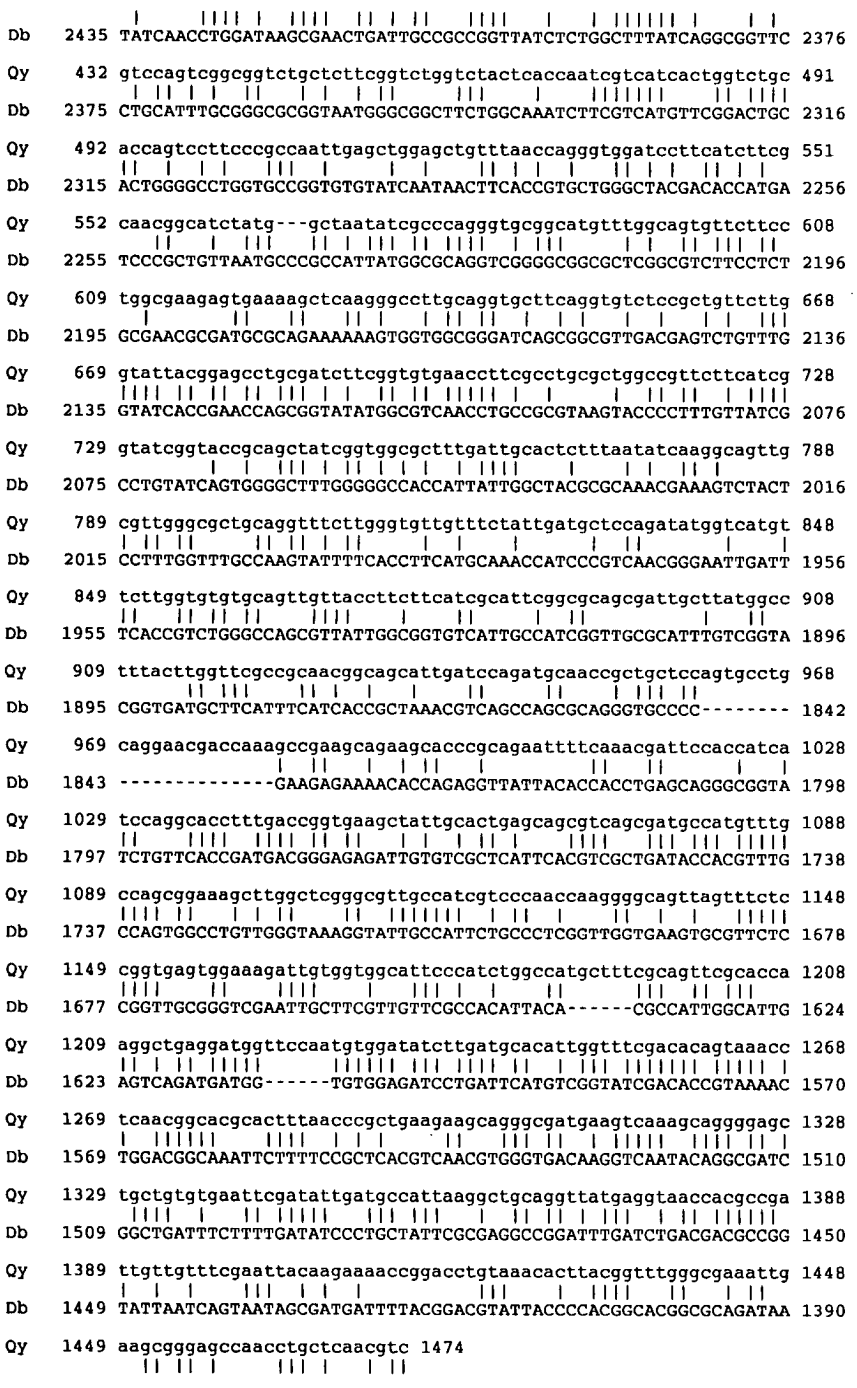
SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	109.6	7.2	3615	1	US-08-920-812-17	Sequence 17, Appl
c	2	109.6	7.2	3615	1	US-08-920-827-17	Sequence 17, Appl
c	3	109.6	7.2	3615	1	US-08-921-177-17	Sequence 17, Appl
c	4	109.6	7.2	3615	1	US-08-362-577C-17	Sequence 17, Appl
c	5	109.6	7.2	3615	2	US-08-920-828-17	Sequence 17, Appl
c	6	97.8	6.4	465	2	US-08-673-190A-3	Sequence 3, Appli
	7	71.6	4.7	357	2	US-08-673-190A-6	Sequence 6, Appli
	8	43.2	2.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	9	39.8	2.6	1314	4	US-08-936-165A-153	Sequence 153, App
c	10	37.2	2.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	11	34.6	2.3	46899	1	US-08-471-119A-1	Sequence 1, Appli
c	12	32.8	2.1	12912	2	US-08-460-751-1	Sequence 1, Appli
c	13	32.8	2.1	14060	3	US-08-658-136-4	Sequence 4, Appli
	14	32.6	2.1	477	4	US-09-135-994-1	Sequence 1, Appli
	15	32.6	2.1	80161	3	US-09-036-987A-1	Sequence 1, Appli
	16	32.6	2.1	80161	4	US-09-370-700-1	Sequence 1, Appli
	17	32.4	2.1	292	4	US-09-117-121-29	Sequence 29, Appl
c	18	32.2	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl
c	19	32	2.1	1008	4	US-09-198-955A-9	Sequence 9, Appli
	20	32	2.1	1901	1	US-08-153-848-43	Sequence 43, Appl
	21	32	2.1	1901	3	US-09-299-843A-43	Sequence 43, Appl
	22	32	2.1	1901	4	US-09-088-337B-43	Sequence 43, Appl
	23	32	2.1	1901	5	PCT-US93-11153-43	Sequence 43, Appl
	24	32	2.1	2232	1	US-08-241-766-12	Sequence 12, Appl
	25	32	2.1	2453	5	PCT-US95-07180-1	Sequence 1, Appli
	26	32	2.1	3120	1	US-08-491-146-1	Sequence 1, Appli
	27	32	2.1	3120	1	US-08-241-766-11	Sequence 11, Appl
	28	32	2.1	3120	1	US-08-234-011-1	Sequence 1, Appli
	29	32	2.1	3120	2	US-08-701-062A-1	Sequence 1, Appli
	30	32	2.1	3533	4	US-08-985-916-15	Sequence 15, Appl
c	31	32	2.1	5163	3	US-08-700-651-1	Sequence 1, Appli
c	32	32	2.1	5163	3	US-08-928-361B-4	Sequence 4, Appli
c	33	32	2.1	5318	3	US-08-700-651-2	Sequence 2, Appli
c	34	32	2.1	5318	3	US-08-928-361B-3	Sequence 3, Appli
	35	32	2.1	10348	2	US-08-457-273B-41	Sequence 41, Appl
	36	32	2.1	10348	3	US-08-556-419-13	Sequence 13, Appl
	37	32	2.1	10348	4	US-09-041-886-14	Sequence 14, Appl
	38	32	2.1	10366	1	US-08-246-982A-5	Sequence 5, Appli
	39	32	2.1	10366	1	US-08-453-265-5	Sequence 5, Appli
c	40	31.8	2.1	1227	3	US-09-074-912-3	Sequence 3, Appli
c	41	31.8	2.1	1227	4	US-09-290-136-3	Sequence 3, Appli
c	42	31.6	2.1	2336	4	US-09-228-986-10	Sequence 10, Appl
c	43	31.6	2.1	5511	3	US-08-928-361B-2	Sequence 2, Appli
c	44	31.6	2.1	7334	3	US-08-928-361B-1	Sequence 1, Appli
	45	31.6	2.1	7655	1	US-08-619-554-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-920-812-17/c
; Sequence 17, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  Clinical Isolate EC-24
US-08-920-812-17
```

```
  Query Match             7.2%;  Score 109.6;  DB 1;  Length 3615;
  Best Local Similarity   46.3%;  Pred. No. 2.3e-23;
  Matches  540;  Conservative    0;  Mismatches  589;  Indels   37;  Gaps    4;

Qy     312 tcactccagtgctgacgttgctgctcaccggattccttacattcatcgccattggcccag 371
           |||| ||| ||||     || || | | |  | | || ||  |     | || ||
Db    2495 TCACACCATTGCTATGTCTGATGGTTATCACACCCGTCACCTTTCTGCTGGTGGGGCCGC 2436

Qy     372 caatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatgatttcggtg 431
           |    |||| |  |||| || | ||   ||||   |   | ||||||  |    | |
Db    2435 TATCAACCTGGATAAGCGAACTGATTGCCGCCGGTTATCTCTGGCTTTATCAGGCGGTTC 2376

Qy     432 gtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcactggtctgc 491
           | || | |  ||   | |  | ||     |||    |    |||||||    || ||||
Db    2375 CTGCATTTGCGGGCGCGGTAATGGGCGGCTTCTGGCAAATCTTCGTCATGTTCGGACTGC 2316

Qy     492 accagtccttcccgccaattgagctggagctgtttaaccagggtggatccttcatcttcg 551
           ||  |  | |   |||  |      |  |    || | |   |    || | |  || | |
Db    2315 ACTGGGGCCTGGTGCCGGTGTGTATCAATAACTTCACCGTGCTGGGCTACGACACCATGA 2256

Qy     552 caacggcatctatg---gctaatatcgcccagggtgcggcatgtttggcagtgttcttcc 608
              ||   |  |||   || | ||| || ||||  | |||     | |  || ||| ||
Db    2255 TCCCGCTGTTAATGCCCGCCATTATGGCGCAGGTCGGGGCGGCGCTCGGCGTCTTCCTCT 2196

Qy     609 tggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttg 668
           |       ||  ||  || |  | || ||  |  | |   |  |   | |  |||
Db    2195 GCGAACGCGATGCGCAGAAAAAAGTGGTGGCGGGATCAGCGGCGTTGACGAGTCTGTTTG 2136

Qy     669 gtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcg 728
           |||| || || || ||| | |  || || ||||| | |      |  || ||  | ||||
Db    2135 GTATCACCGAACCAGCGGTATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076

Qy     729 gtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaaggcagttg 788
           |  |  ||| | || | |  | ||||    |    | |  || |
Db    2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCATTATTGGCTACGCGCAAACGAAAGTCTACT 2016

Qy     789 cgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatatggtcatgt 848
           | || ||     || || | ||     |  |    |      |  ||       |    |
Db    2015 CCTTTGGTTTGCCAAGTATTTTCACCTTCATGCAAACCATCCCGTCAACGGGAATTGATT 1956

Qy     849 tcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgcttatggcc 908
           ||  || || ||   ||||    |    ||     |   ||          |   ||
Db    1955 TCACCGTCTGGGCCAGCGTTATTGGCGGTGTCATTGCCATCGGTTGCGCATTTGTCGGTA 1896

Qy     909 tttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctg 968
             || |||   || |  |   |    ||     ||    | ||| ||
Db    1895 CGGTGATGCTTCATTTCATCACCGCTAAACGTCAGCCAGCGCAGGGTGCCCC-------- 1842

Qy     969 caggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattccaccatca 1028
                   |  ||   |  | ||   |          ||   ||       |   |
Db    1843 --------------GAAGAGAAAACACCAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798

Qy    1029 tccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttg 1088
           ||    ||||  |||| || ||   | |  | || |    ||||   ||| ||| |||||
Db    1797 TCTGTTCACCGATGACGGGAGAGATTGTGTCGCTCATTCACGTCGCTGATACCACGTTTG 1738

Qy    1089 ccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctc 1148
           |||| ||   | | ||    ||  |||||||  | ||  |    ||  |  |   |||||
Db    1737 CCAGTGGCCTGTTGGGTAAAGGTATTGCCATTCTGCCCTCGGTTGGTGAAGTGCGTTCTC 1678

Qy    1149 cggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagttcgcacca 1208
           ||||    ||    ||||    |   ||| |  |    ||      |||  || |||
Db    1677 CGGTTGCGGGTCGAATTGCTTCGTTGTTCGCCACATTACA------CGCCATTGGCATTG 1624

Qy    1209 aggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacc 1268
           || | || |||||      |||||| ||| |||| ||  | ||| ||||||| ||||| |
Db    1623 AGTCAGATGATGG------TGTGGAGATCCTGATTCATGTCGGTATCGACACCGTAAAAC 1570

Qy    1269 tcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagc 1328
           |  ||||||    ||||  | | |     ||   ||| ||  | |||||  |||| || |
Db    1569 TGGACGGCAAATTCTTTTCCGCTCACGTCAACGTGGGTGACAAGGTCAATACAGGCGATC 1510

Qy    1329 tgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaaccacgccga 1388
           |||| |   || ||||||  ||| |||   |  || || | |||  | || ||||||
Db    1509 GGCTGATTTCTTTTGATATCCCTGCTATTCGCGAGGCCGGATTTGATCTGACGACGCCGG 1450

Qy    1389 ttgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattg 1448
           |  |  |    ||| | | |          |||    |  ||||   ||   | ||
Db    1449 TATTAATCAGTAATAGCGATGATTTTACGGACGTATTACCCCACGGCACGGCGCAGATAA 1390

Qy    1449 aagcgggagccaacctgctcaacgtc 1474
             || || |     ||| |   | ||
Db    1389 GCGCAGGTGAACCGCTGTTATCCATC 1364


RESULT   2
US-08-920-827-17/c
; Sequence 17, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  Clinical Isolate EC-24
US-08-920-827-17


  Query Match             7.2%;  Score 109.6;  DB 1;  Length 3615;
  Best Local Similarity   46.3%;  Pred. No. 2.3e-23;
  Matches  540;  Conservative    0;  Mismatches  589;  Indels   37;  Gaps    4;

Qy     312 tcactccagtgctgacgttgctgctcaccggattccttacattcatcgccattggcccag 371
           |||| ||| ||||     || || | | |  | | || ||  |     | || ||
Db    2495 TCACACCATTGCTATGTCTGATGGTTATCACACCCGTCACCTTTCTGCTGGTGGGGCCGC 2436

Qy     372 caatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatgatttcggtg 431
```



Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 3
US-08-921-177-17/c
; Sequence 17, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
US-08-921-177-17

Query Match 7.2%; Score 109.6; DB 1; Length 3615;
Best Local Similarity 46.3%; Pred. No. 2.3e-23;
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;

```
Qy    552 caacggcatctatg---gctaatatcgcccagggtgcggcatgtttggcagtgttcttcc 608
          ||    |  |||   || | ||| || ||||  | |||     | |  || ||| ||
Db   2255 TCCCGCTGTTAATGCCCGCCATTATGGCGCAGGTCGGGGCGGCGCTCGGCGTCTTCCTCT 2196

Qy    609 tggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttg 668
          |       ||   ||   || |   | || ||  |  | |   |  |   | |  |||
Db   2195 GCGAACGCGATGCGCAGAAAAAAGTGGTGGCGGGATCAGCGGCGTTGACGAGTCTGTTTG 2136

Qy    669 gtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcg 728
          |||| || || || ||| | |  || || ||||| |  |     |  || ||  | ||||
Db   2135 GTATCACCGAACCAGCGGTATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076

Qy    729 gtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaaggcagttg 788
              | |  ||| | || | | |  | ||||    |    | |  || |
Db   2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCATTATTGGCTACGCGCAAACGAAAGTCTACT 2016

Qy    789 cgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatatggtcatgt 848
          | || ||     || || | ||    |  |     |    |  ||      |   |
Db   2015 CCTTTGGTTTGCCAAGTATTTTCACCTTCATGCAAACCATCCCGTCAACGGGAATTGATT 1956

Qy    849 tcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgcttatggcc 908
          ||   || || ||    ||||    |    ||     |  ||          |   ||
Db   1955 TCACCGTCTGGGCCAGCGTTATTGGCGGTGTCATTGCCATCGGTTGCGCATTTGTCGGTA 1896

Qy    909 tttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctg 968
              || |||    || |  |   |    ||     ||    | ||| ||
Db   1895 CGGTGATGCTTCATTTCATCACCGCTAAACGTCAGCCAGCGCAGGGTGCCCC-------- 1842

Qy    969 caggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattccaccatca 1028
                    |  ||   |  | ||  |          ||  ||      |   |
Db   1843 -------------GAAGAGAAAACACCAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798

Qy   1029 tccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttg 1088
          ||    ||||  |||| || ||   | |  | || |    ||||   ||| ||| |||||
Db   1797 TCTGTTCACCGATGACGGGAGAGATTGTGTCGCTCATTCACGTCGCTGATACCACGTTTG 1738

Qy   1089 ccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctc 1148
          |||| ||   | | ||    ||  |||||||  | ||  |    ||  |  |   |||||
Db   1737 CCAGTGGCCTGTTGGGTAAAGGTATTGCCATTCTGCCCTCGGTTGGTGAAGTGCGTTCTC 1678

Qy   1149 cggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagttcgcacca 1208
          ||||    ||   ||||    |  ||| |  |    ||      ||| || |||
Db   1677 CGGTTGCGGGTCGAATTGCTTCGTTGTTCGCCACATTACA------CGCCATTGGCATTG 1624

Qy   1209 aggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacc 1268
          || | || |||||      |||||| ||| |||| ||  | ||| |||||||| ||||| |
Db   1623 AGTCAGATGATGG------TGTGGAGATCCTGATTCATGTCGGTATCGACACCGTAAAAC 1570

Qy   1269 tcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagc 1328
          |  ||||||    ||||  | | |     ||   ||| ||  | |||||  |||| || |
Db   1569 TGGACGGCAAATTCTTTTCCGCTCACGTCAACGTGGGTGACAAGGTCAATACAGGCGATC 1510

Qy   1329 tgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaaccacgccga 1388
           |||| |   || |||||   ||| |||   |  || || | |||  | || ||||||
Db   1509 GGCTGATTTCTTTTGATATCCCTGCTATTCGCGAGGCCGGATTTGATCTGACGACGCCGG 1450

Qy   1389 ttgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattg 1448
          | | |   |||  | | |         |||    |  ||||   ||   | ||
Db   1449 TATTAATCAGTAATAGCGATGATTTTACGGACGTATTACCCCACGGCACGGCGCAGATAA 1390

Qy   1449 aagcgggagccaacctgctcaacgtc 1474
            || || |     ||| |   | ||
Db   1389 GCGCAGGTGAACCGCTGTTATCCATC 1364


RESULT   4
US-08-362-577C-17/c
; Sequence 17, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  Clinical Isolate EC-24
US-08-362-577C-17


  Query Match             7.2%;  Score 109.6;  DB 1;  Length 3615;
  Best Local Similarity   46.3%;  Pred. No. 2.3e-23;
  Matches  540;  Conservative    0;  Mismatches  589;  Indels   37;  Gaps    4;

Qy    312 tcactccagtgctgacgttgctgctcaccggattccttacattcatcgccattggcccag 371
          |||| ||| ||||     || || | | |  |  | | | || ||  |     | || ||
Db   2495 TCACACCATTGCTATGTCTGATGGTTATCACACCCGTCACCTTTCTGCTGGTGGGGCCGC 2436

Qy    372 caatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatgatttcggtg 431
          |     |||| |  |||  || | ||   ||||    |   | ||||| |     | |
Db   2435 TATCAACCTGGATAAGCGAACTGATTGCCGCCGGTTATCTCTGGCTTTATCAGGCGGTTC 2376

Qy    432 gtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcactggtctgc 491
          | || | |  ||   | |  | ||     |||     |    |||||||    || ||||
Db   2375 CTGCATTTGCGGGCGCGGTAATGGGCGGCTTCTGGCAAATCTTCGTCATGTTCGGACTGC 2316

Qy    492 accagtccttcccgccaattgagctggagctgtttaaccagggtggatccttcatcttcg 551
          ||  |  | |  |||  |     | |    || | |  |   || | |  || | | |
Db   2315 ACTGGGGCCTGGTGCCGGTGTGTATCAATAACTTCACCGTGCTGGGCTACGACACCATGA 2256

Qy    552 caacggcatctatg---gctaatatcgcccagggtgcggcatgtttggcagtgttcttcc 608
            ||    |  |||   || | ||| || ||||  | |||     | |  || ||| ||
Db   2255 TCCCGCTGTTAATGCCCGCCATTATGGCGCAGGTCGGGGCGGCGCTCGGCGTCTTCCTCT 2196

Qy    609 tggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttg 668
          |       ||   ||   || |   | || ||  |  | |   |  |   | |  |||
Db   2195 GCGAACGCGATGCGCAGAAAAAAGTGGTGGCGGGATCAGCGGCGTTGACGAGTCTGTTTG 2136

Qy    669 gtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcg 728
          |||| || || || ||| | |  || || ||||| |  |     |  || ||  | ||||
Db   2135 GTATCACCGAACCAGCGGTATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076
```

```
Qy    729 gtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaaggcagttg 788
Db   2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCATTATTGGCTACGCGCAAACGAAAGTCTACT 2016

Qy    789 cgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatatggtcatgt 848
Db   2015 CCTTTGGTTTGCCAAGTATTTTCACCTTCATGCAAACCATCCCGTCAACGGGAATTGATT 1956

Qy    849 tcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgcttatggcc 908
Db   1955 TCACCGTCTGGGCCAGCGTTATTGGCGGTGTCATTGCCATCGGTTGCGCATTTGTCGGTA 1896

Qy    909 tttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctg 968
Db   1895 CGGTGATGCTTCATTTCATCACCGCTAAACGTCAGCCAGCGCAGGGTGCCCC-------- 1842

Qy    969 caggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattccaccatca 1028
Db   1843 -------------GAAGAGAAAACACCAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798

Qy   1029 tccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttg 1088
Db   1797 TCTGTTCACCGATGACGGGAGAGATTGTGTCGCTCATTCACGTCGCTGATACCACGTTTG 1738

Qy   1089 ccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctc 1148
Db   1737 CCAGTGGCCTGTTGGGTAAAGGTATTGCCATTCTGCCCTCGGTTGGTGAAGTGCGTTCTC 1678

Qy   1149 cggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagttcgcacca 1208
Db   1677 CGGTTGCGGGTCGAATTGCTTCGTTGTTCGCCACATTACA------CGCCATTGGCATTG 1624

Qy   1209 aggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacc 1268
Db   1623 AGTCAGATGATGG------TGTGGAGATCCTGATTCATGTCGGTATCGACACCGTAAAAC 1570

Qy   1269 tcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagc 1328
Db   1569 TGGACGGCAAATTCTTTTCCGCTCACGTCAACGTGGGTGACAAGGTCAATACAGGCGATC 1510

Qy   1329 tgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaaccacgccga 1388
Db   1509 GGCTGATTTCTTTTGATATCCCTGCTATTCGCGAGGCCGGATTTGATCTGACGACGCCGG 1450

Qy   1389 ttgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattg 1448
Db   1449 TATTAATCAGTAATAGCGATGATTTTACGGACGTATTACCCCACGGCACGGCGCAGATAA 1390

Qy   1449 aagcgggagccaacctgctcaacgtc 1474
Db   1389 GCGCAGGTGAACCGCTGTTATCCATC 1364
```

```
RESULT   5
US-08-920-828-17/c
; Sequence 17, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3615 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  Clinical Isolate EC-24
US-08-920-828-17


  Query Match             7.2%;  Score 109.6;  DB 2;  Length 3615;
  Best Local Similarity   46.3%;  Pred. No. 2.3e-23;
  Matches  540;  Conservative    0;  Mismatches  589;  Indels   37;  Gaps    4;

Qy    312 tcactccagtgctgacgttgctgctcaccggattccttacattcatcgccattggcccag 371
Db   2495 TCACACCATTGCTATGTCTGATGGTTATCACACCCGTCACCTTTCTGCTGGTGGGGCCGC 2436

Qy    372 caatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatgatttcggtg 431
Db   2435 TATCAACCTGGATAAGCGAACTGATTGCCGCCGGTTATCTCTGGCTTTATCAGGCGGTTC 2376

Qy    432 gtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcactggtctgc 491
Db   2375 CTGCATTTGCGGGCGCGGTAATGGGCGGCTTCTGGCAAATCTTCGTCATGTTCGGACTGC 2316

Qy    492 accagtccttcccgccaattgagctggagctgtttaaccagggtggatccttcatcttcg 551
Db   2315 ACTGGGGCCTGGTGCCGGTGTGTATCAATAACTTCACCGTGCTGGGCTACGACACCATGA 2256

Qy    552 caacggcatctatg---gctaatatcgcccagggtgcggcatgtttggcagtgttcttcc 608
Db   2255 TCCCGCTGTTAATGCCCGCCATTATGGCGCAGGTCGGGGCGGCGCTCGGCGTCTTCCTCT 2196

Qy    609 tggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttg 668
Db   2195 GCGAACGCGATGCGCAGAAAAAAGTGGTGGCGGGATCAGCGGCGTTGACGAGTCTGTTTG 2136

Qy    669 gtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcg 728
Db   2135 GTATCACCGAACCAGCGGTATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076

Qy    729 gtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaaggcagttg 788
Db   2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCATTATTGGCTACGCGCAAACGAAAGTCTACT 2016

Qy    789 cgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatatggtcatgt 848
Db   2015 CCTTTGGTTTGCCAAGTATTTTCACCTTCATGCAAACCATCCCGTCAACGGGAATTGATT 1956

Qy    849 tcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgcttatggcc 908
```

```
Db   1955 TCACCGTCTGGGCCAGCGTTATTGGCGGTGTCATTGCCATCGGTTGCGCATTTGTCGGTA 1896

Qy    909 tttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctg 968
Db   1895 CGGTGATGCTTCATTTCATCACCGCTAAACGTCAGCCAGCGCAGGGTGCCCC-------- 1842

Qy    969 caggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattccaccatca 1028
Db   1843 ----------GAAGAGAAAACACCAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798

Qy   1029 tccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttg 1088
Db   1797 TCTGTTCACCGATGACGGGAGAGATTGTGTCGCTCATTCACGTCGCTGATACCACGTTTG 1738

Qy   1089 ccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctc 1148
Db   1737 CCAGTGGCCTGTTGGGTAAAGGTATTGCCATTCTGCCCTCGGTTGGTGAAGTGCGTTCTC 1678

Qy   1149 cggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagttcgcacca 1208
Db   1677 CGGTTGCGGGTCGAATTGCTTCGTTGTTCGCCACATTACA------CGCCATTGGCATTG 1624

Qy   1209 aggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacc 1268
Db   1623 AGTCAGATGATGG------TGTGGAGATCCTGATTCATGTCGGTATCGACACCGTAAAAC 1570

Qy   1269 tcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagc 1328
Db   1569 TGGACGGCAAATTCTTTTCCGCTCACGTCAACGTGGGTGACAAGGTCAATACAGGCGATC 1510

Qy   1329 tgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaaccacgccga 1388
Db   1509 GGCTGATTTCTTTTGATATCCCTGCTATTCGCGAGGCCGGATTTGATCTGACGACGCCGG 1450

Qy   1389 ttgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattg 1448
Db   1449 TATTAATCAGTAATAGCGATGATTTTACGGACGTATTACCCCACGGCACGGCGCAGATAA 1390

Qy   1449 aagcgggagccaacctgctcaacgtc 1474
Db   1389 GCGCAGGTGAACCGCTGTTATCCATC 1364
```

```
RESULT   6
US-08-673-190A-3/c
; Sequence 3, Application US/08673190A
; Patent No. 5985668
;  GENERAL INFORMATION:
;    APPLICANT:  Mattes, Ralf
;    APPLICANT:  Klein, Kathrin
;    APPLICANT:  Stegmaier, Sabine
;    TITLE OF INVENTION:  Sucrose Metabolism Mutants.
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,190A
;      FILING DATE:  27-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Forman, David S.
;      REGISTRATION NUMBER:  33,694
;      REFERENCE/DOCKET NUMBER:  06473.0001-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)408-4000
;      TELEFAX:  (202)408-4400
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  465 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Protaminobacter rubrum
US-08-673-190A-3


  Query Match             6.4%;  Score 97.8;  DB 2;  Length 465;
  Best Local Similarity   53.5%;  Pred. No. 2.8e-20;
  Matches  204;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    145 cgccaccatggctgcgggcgaaatgccaatgtggtccctgtttggtttagatgttgccca 204
Db    441 CGCCTGGGGCGTCGCGGGCGGTTTCCACACCATGAATTTCTTCGGACTCGATATCGCCAT 382

Qy    205 agccggttaccagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgat 264
Db    381 GATCGGCTATCAGGGTACGGTGTTCCCGGTGCTGCTGACGGTGTGGTTTATGAGCGTGCT 322

Qy    265 cgagaagttcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgct 324
Db    321 GGAAAAACAGCTGCGCAGAGTGATCCCCAACGCGCTGGACCTGATCCTGACGCCATTCCT 262

Qy    325 gacgttgctgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggt 384
Db    261 GACGGTGGTGATCTCCGGCTTCGTCGCCATGCTGTTTATCGGCCCCGCCGGGCGCACACT 202

Qy    385 gggcgatgtgctggcacacggtctacagggactttatgatttcggtggtccagtcggcgg 444
Db    201 GGGTGACGGCATCTCCTTCGTGCTCAGCACGCTGATTGCCCACGCCGGTTGGTTCGCCGG 142

Qy    445 tctgctcttcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttccc 504
Db    141 ATTCCTGTTCGGCGGCCTGTATTCCGCCATCGTCATCACCGGCATTCACCACAGCTTCCA 82

Qy    505 gccaattgagctggagctgtt 525
Db     81 CGCCGTAGAGGGGGGGCTGCT 61
```

```
RESULT   7
US-08-673-190A-6
; Sequence 6, Application US/08673190A
; Patent No. 5985668
;  GENERAL INFORMATION:
;    APPLICANT:  Mattes, Ralf
;    APPLICANT:  Klein, Kathrin
;    APPLICANT:  Stegmaier, Sabine
;    TITLE OF INVENTION:  Sucrose Metabolism Mutants.
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/673,190A
;      FILING DATE:  27-JUN-1996
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Forman, David S.
;      REGISTRATION NUMBER:  33,694
;      REFERENCE/DOCKET NUMBER:  06473.0001-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)408-4000
;      TELEFAX:  (202)408-4400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  357 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterobacter species
US-08-673-190A-6


  Query Match             4.7%;  Score 71.6;  DB 2;  Length 357;
  Best Local Similarity   53.1%;  Pred. No. 2.6e-12;
  Matches  152;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy    183 tgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtgg 242
          | || ||  | || || ||      ||| |||||||||||||  | || ||||||  ||| 
Db     67 TCTTCGGCATCGAAGTGGCGATGATCGGCTACCAGGGCACCGTCTTCCCGGTGCTGCTGG 126

Qy    243 tttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcag 302
              ||| || ||   | | |||||||    |||| |       ||  | | |     |
Db    127 CGGTGTGGTTTATGAGCATGGTCGAGAAACGGCTGCGCSSCGTTATCCCTGACGCGCTGG 186

Qy    303 acttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcca 362
          || |  |  |||| ||  | ||||||  || |  || |||| ||  |  |  |  |    |
Db    187 ACCTGATCCTCACCCCGTTCCTGACGGTGATTATCTCCGGCTTTATCGCCCTGCTGCTGA 246

Qy    363 ttggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatg 422
          | ||||| ||    |||  | | |||||  |   |  |       ||       |||
Db    247 TCGGCCCGGCCGGTCGCGCGCTCGGCGACGGTATTTCGTTTATCCTCAGCACGCTTATCA 306

Qy    423 atttcggtggtccagtcggcggtctgctcttcggtctggtctactc 468
              ||  ||     | |  || ||||| |||||     |||||||
Db    307 GCCACGCCGGCTGGCTGGCGGGCCTGCTGTTCGGCGGCCTCTACTC 352


RESULT   8
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             2.8%;  Score 43.2;  DB 1;  Length 7218;
  Best Local Similarity   7.1%;  Pred. No. 0.0078;
  Matches   30;  Conservative  207;  Mismatches  185;  Indels    0;  Gaps    0;

Qy    147 ccaccatggctgcgggcgaaatgccaatgtggtccctgtttggtttagatgttgcccaag 206
          | | | |||||||  ||    |  |     || | |:::: :::  :::   : :: :::
Db   1034 CGAGCTTGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYY 1093

Qy    207 ccggttaccagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgatcg 266
          ::  :: ::    : :: : :::::: : :: :  ::::::   ::::  :  :  ::
Db   1094 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1153

Qy    267 agaagttcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgctga 326
              ::::: : :   :  :::     : :: :   ::::::  :: ::::  : ::
Db   1154 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1213

Qy    327 cgttgctgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggtgg 386
          : :: :: ::: ::   :::::: : ::: :: :: ::  :::  :  : : ::   :
Db   1214 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1273

Qy    387 gcgatgtgctggcacacggtctacagggactttatgatttcggtggtccagtcggcggtc 446
          :  : : ::  : : :  ::: :    :::: :  :::: :  ::: :: :  : ::
Db   1274 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1333

Qy    447 tgctcttcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttcccgc 506
          : :::::: :::  ::: ::: ::  :: :: :: ::  ::: : ::  :::::::: :
Db   1334 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1393

Qy    507 caattgagctggagctgtttaaccagggtggatccttcatcttcgcaacggcatctatgg 566
          :  ::   ::    :: :::  ::    :   :::::: :: |  |||   | |||||
Db   1394 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCT 1453

Qy    567 ct 568
          ||
Db   1454 CT 1455


RESULT   9
US-08-936-165A-153
; Sequence 153, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
```

```
;    APPLICANT:  Ward, Judith
;    TITLE OF INVENTION:  No. 6348582el Prokaryotic Polynucleotides,
;    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  153:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-936-165A-153


  Query Match           2.6%;  Score 39.8;  DB 4;  Length 1314;
  Best Local Similarity   48.9%;  Pred. No. 0.031;
  Matches  107;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy     153 tggctgcgggcgaaatgccaatgtggtccctgtttggtttagatgttgcccaagccggtt 212
           ||||    ||  | || || | ||||  | | ||||| |||||  ||   ||      ||
Db    1045 TGGCAAAAGGGAATATTCCGACGTGGAACTTATTTGGCTTAGAGATTAAGCAGTTGAATT 1104

Qy     213 accagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgatcgagaagt 272
           |||| ||    ||| | |||||  |  | |   |||   |||| || |  ||    ||
Db    1105 ACCAAGGTCAAGTGTTGCCTGTTTTAATTGCAGCTTATGTTCTAGCTAAAATTTGAAAAA 1164

Qy     273 tcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgctgacgttgc 332
             |  |  | | |      |    |  |     |       |       | ||| |
Db    1165 GGATTAAATAAAGTCGTTCACGATTCGATAAAAATGTTGGTCGTTGGACCGTAACGCTTT 1224

Qy     333 tgctcaccggattccttacattcatcgccattggcccag 371
           |  | || |||||  |  |||| ||   |||||| ||||
Db    1225 TAGTTACTGGATTTTTAGCATTTATTATCATTGGACCAG 1263


RESULT  10
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match           2.4%;  Score 37.2;  DB 4;  Length 4403765;
  Best Local Similarity   43.7%;  Pred. No. 19;
  Matches  165;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy     627 tcaagggccttgcaggtgcttcaggtgtctccgctgttcttggtattacggagcctgcga 686
           | |  | | | || |  | | |    ||| |||| | |       | |      | |
Db 3923279 TTACCGCCGTCGCCGCCGTTGCCCCCGTCGCCGCCGATGCCCTGGCTGCCAGCGTTACCG 3923220

Qy     687 tcttcggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagcta 746
           |    ||| |  |   ||||   || || |       |  ||   | | | |||  || |
Db 3923219 CCGGTAGTGGGGCCGGGCGCCCCGGCGGGGCCGGGCGCGCCGCCGTTGCTTCCGGTGCCA 3923160

Qy     747 tcggtggcgctttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtt 806
           || || | |       ||  |      |    | || | ||    || ||     ||
Db 3923159 CCGTTGCCACCGGCCCCGCCGTTGGGTGTGGGCAAGCTGGTGGCCGGGACGAACTGGGCG 3923100

Qy     807 tcttgggtgttgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagttg 866
           |  ||||  ||       | | |  | |  |  | | |   |  ||    || ||||
Db 3923099 CCGACGGTGCCGTGGGCGCCGTTACCGCCGTCACCGCCGTTGCCGCCGTTGCCGCCGTTG 3923040

Qy     867 ttaccttcttcatcgcattcggcgcagcgattgcttatggcctttacttggttcgccgca 926
           ||| |   |  |||    | ||  ||    | ||  | ||    |   | |  ||||
Db 3923039 CCACCGTTGCCGCCGCTAAGGCCGGTGCCGCCGGTTCCGCCCGCACCGCCGCTACCCGCG 3922980

Qy     927 acggcagcattgatccagatgcaaccgctgctccagtgcctgcaggaacgaccaaagccg 986
           | | || |||   | | |||  |||||||   ||   | || ||    |      |||
Db 3922979 GTGCCGGCGTTGCCGCCGTTGCCGCCGCTGCCGGAGGCGCCGCTGGTGGTATTGCCACCG 3922920

Qy     987 aagcagaagcacccgcag 1004
           |||  |  | | |||| |
Db 3922919 AAGTCGCCGGAGCCGCCG 3922902


RESULT  11
US-08-471-119A-1/c
; Sequence 1, Application US/08471119A
; Patent No. 5827706
;  GENERAL INFORMATION:
;    APPLICANT:  Leitner, Ernst
;    APPLICANT:  Schneider, Elisabeth
;    APPLICANT:  Schoergendorfer, Kurt
;    APPLICANT:  Weber, Gerhard
;    TITLE OF INVENTION:  Cyclosporin Synthetase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5827706artis Corporation
;      STREET:  59 Route 10
;      CITY:  East Hanover
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07936
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/471,119A
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kassenoff, Melvyn
;       REGISTRATION NUMBER:  26,389
;       REFERENCE/DOCKET NUMBER:  100-8029/CONT/CONT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201 503 8474
;       TELEFAX:  201 503 8807
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  46899 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Tolypocladium niveum
;       STRAIN:  ATCC 34921
US-08-471-119A-1


  Query Match             2.3%;  Score 34.6;  DB 1;  Length 46899;
  Best Local Similarity   55.4%;  Pred. No. 10;
  Matches   67;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
           || | ||||||| |||| ||   || |  ||   || ||  ||     |  |||  |||
Db  29376 CGATACGGCGGCTCTGAGCTCGCTGGAGTGGCCGGTTCGTCAGGGTTCGAAGTTGTCGAC 29317

Qy     121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
           |||| ||  |||    || |||  |     || | |||||  || |||    ||  | |
Db  29316 CTTGATGGTCGGTAGGGAAGTGTACCAGGACACGACTGCGCTCGCAATTTGCATCGGTTG 29257

Qy     181 c 181
           |
Db  29256 C 29256


RESULT  12
US-08-460-751-1/c
; Sequence 1, Application US/08460751
; Patent No. 5891628
;  GENERAL INFORMATION:
;    APPLICANT:  Reeders, Stephen
;    APPLICANT:  Schneider, Michael
;    APPLICANT:  Glucksmann, Sandra
;    TITLE OF INVENTION:  IDENTIFICATION OF POLYCYSTIC KIDNEY
;    TITLE OF INVENTION:  DISEASE GENE, DIAGNOSTICS AND TREATMENT
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,751
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/413,580
;      FILING DATE:  03-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7638-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12912 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..12912
US-08-460-751-1


  Query Match             2.1%;  Score 32.8;  DB 2;  Length 12912;
  Best Local Similarity   48.9%;  Pred. No. 17;
  Matches   88;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy     235 gctggtggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaaggg 294
           | ||||||  || |   | | | |  |  || | | | ||  ||  |||     |
Db    7059 GTTGGTGGCCTCCTCCTTGCGGCCGGCCTTCCACACGGTCAGGCTGAAGGTGTACTCCAC 7000

Qy     295 cactgcagacttcctgatcactccagtgctgacgttgctgctcaccggattccttacatt 354
           | || | |  ||    |  |  | || ||||  |||||||| |  |   ||  |  ||
Db    6999 GCCAGCCGCCAGCCGCTCCCGTGGAATGGTGACCGTGCTGCTCCCGCGGGGCCCAAAGTT 6940

Qy     355 catcgccattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacaggg 414
           || |||       |||||    | ||| || |  |     | ||| ||  |   || | |
Db    6939 CAGCGCACACCCGCCAGCCTCCCTCTGTGTCGAAGCCACACAGGCCCAGTGGAAACTGAG 6880


RESULT  13
US-08-658-136-4/c
; Sequence 4, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
```

```
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14060 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  135..13040
US-08-658-136-4


  Query Match             2.1%;  Score 32.8;  DB 3;  Length 14060;
  Best Local Similarity   48.9%;  Pred. No. 18;
  Matches   88;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy    235 gctggtggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaaggg 294
          | ||||||  || |    | | | |   |  || | | | ||  ||  |||     |
Db   7193 GTTGGTGGCCTCCTCCTTGCGGCCGGCCTTCCACACGGTCAGGCTGAAGGTGTACTCCAC 7134

Qy    295 cactgcagacttcctgatcactccagtgctgacgttgctgctcaccggattccttacatt 354
            | ||  | |  ||    |  |  | || ||||  |||||||| |  |   ||  |  ||
Db   7133 GCCAGCCGCCAGCCGCTCCCGTGGAATGGTGACCGTGCTGCTCCCGCGGGGCCCAAAGTT 7074

Qy    355 catcgccattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacaggg 414
          ||  |||       |||||    | ||| || |  |     | ||| ||  |   || | |
Db   7073 CAGCGCACACCCGCCAGCCTCCCTCTGTGTCGAAGCCACACAGGCCCAGTGGAAACTGAG 7014


RESULT  14
US-09-135-994-1
; Sequence 1, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
;  APPLICANT: Ranum et al.
;  TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
;  FILE REFERENCE: University of Minnesota
;  CURRENT APPLICATION NUMBER: US/09/135,994A
;  CURRENT FILING DATE: 1998-08-18
;  EARLIER APPLICATION NUMBER: 60/056,170
;  EARLIER FILING DATE: 1997-08-19
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-135-994-1


  Query Match             2.1%;  Score 32.6;  DB 4;  Length 477;
  Best Local Similarity   60.9%;  Pred. No. 2.7;
  Matches   53;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    920 cgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctgcaggaacgacc 979
          ||  ||| |||| ||   |   |||  ||| | || ||  |||   | ||||  | |  |
Db    108 cggagcagcggccgcccggcagcagcagcagcagcagcagcagcagcagcagcagca 167

Qy    980 aaagccgaagcagaagcacccgcagaa 1006
              ||| | ||||| |||| | |||| |
Db    168 gcagcagcagcagcagcagcagca 194


RESULT  15
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1


  Query Match             2.1%;  Score 32.6;  DB 3;  Length 80161;
  Best Local Similarity   53.5%;  Pred. No. 57;
  Matches   68;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          || ||| |||   | |||   || | ||    | |  |||  |    ||| |   |||||
Db  14636 CGGTTCCGCGCGGAAGAGGATCTCGTCGATCAGCACGGGTGCGATCCTGGCGAGTCCGAG 14695

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
           | ||||  | | |  | |  |  |||| |  ||| | |||  |  | |   ||| | ||
Db  14696 TTCGGTGGTCAGGTCGGCCAGGCTCGCCGCACTGGATCCGGCGGCGAGGATGATGCGTTC 14755

Qy    181 cctgttt 187
            |  | ||
Db  14756 CACGGTT 14762


Search completed: September 25, 2002, 08:47:50
Job time: 9626 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:16:59 ; Search time 234.58 Seconds
(without alignments)
11176.265 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 ctcatggcatctgcgccgtt.....gttgaaaccttgagtgttcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1527	100.0	1527	22	AAF31528	C.glutamicum phosp
c	2	1515	99.2	349980	22	AAH68533	C glutamicum codin
	3	1492	97.7	1983	22	AAH67869	C glutamicum codin
	4	1412.6	92.5	5969	22	AAF32543	Brevibacterium lac
	5	1107.4	72.5	1109	22	AAF31529	C.glutamicum phosp
c	6	190.6	12.5	2913	22	AAH54100	S. epidermidis gen
	7	151.2	9.9	5840	20	AAX12968	Enterococcus faeca
	8	132.2	8.7	2944528	24	ABA03041	Listeria monocytog
c	9	126.6	8.3	3895	19	AAV52334	Streptococcus pneu
	10	113.4	7.4	1884	23	AAS55732	Streptococcus pneu
c	11	108	7.1	3615	15	AAQ55752	Escherichia coli g
c	12	107.8	7.1	9769	19	AAV52163	Streptococcus pneu
c	13	97.8	6.4	465	18	AAT47502	Partial P. rubrum
	14	86.2	5.6	30246	18	AAV74367	Staphylococcus aur
c	15	81.6	5.3	2944528	24	ABA03041	Listeria monocytog
	16	76.4	5.0	465	18	AAV75628	Staphylococcus aur
	17	74.2	4.9	474	20	AAX13576	Enterococcus faeca
	18	71.6	4.7	357	18	AAT47505	Partial Enterobact
c	19	70.4	4.6	9797	20	AAX13487	Enterococcus faeca
	20	67	4.4	8494	19	AAV52296	Streptococcus pneu
	21	66	4.3	1947	22	AAH81338	Escherichia coli p
	22	66	4.3	1947	23	AAS52355	E. coli DNA for ce
c	23	65.2	4.3	350	23	AAS87225	DNA encoding novel
	24	61.6	4.0	1319	23	AAS52023	Staphylococcus aur
	25	61.6	4.0	1455	23	AAS54537	Staphylococcus aur
	26	61.2	4.0	22934	23	AAS59613	Propionibacterium
	27	60.4	4.0	2550	22	AAH54979	S. epidermidis gen
c	28	60	3.9	6285	20	AAX13352	Enterococcus faeca
	29	59.2	3.9	2049	22	AAH68426	C glutamicum codin
	30	59.2	3.9	2172	22	AAF31542	C.glutamicum phosp
	31	59.2	3.9	349980	22	AAH68528	C glutamicum codin
	32	58.4	3.8	1368	23	AAS56059	Salmonella typhi D
c	33	56.8	3.7	29555	18	AAV74517	Staphylococcus aur
	34	54.2	3.5	428	22	AAF31544	C.glutamicum phosp
c	35	53.2	3.5	2996	22	AAH54445	S. epidermidis gen
c	36	53.2	3.5	3081	22	AAH54946	S. epidermidis gen
	37	53.2	3.5	3932	22	AAH54056	S. epidermidis gen
c	38	51.8	3.4	7156	20	AAX12966	Enterococcus faeca
	39	51.4	3.4	1906	20	AAX13595	Enterococcus faeca
c	40	51.4	3.4	28882	19	AAV52273	Streptococcus pneu
	41	51	3.3	2037	23	AAS52021	Staphylococcus aur
	42	51	3.3	2046	23	AAS54535	Staphylococcus aur
	43	51	3.3	29555	18	AAV74517	Staphylococcus aur
c	44	49.2	3.2	796	18	AAV74736	Staphylococcus aur
	45	46.8	3.1	2529	23	AAS54215	Pseudomonas aerugi

ALIGNMENTS

RESULT 1
AAF31528
ID AAF31528 standard; DNA; 1527 BP.
XX
AC AAF31528;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate DNA #1.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-080989/09.
XX
PT Corynebacteium glutamicum nucleic acids encoding phosphoenolpyruvate:

```
PT   sugar phosphotransferase system proteins or their portions, useful for
PT   typing or identifying C. glutamicum or related bacteria, and as markers
PT   for transformation -
XX
PS   Claim 3; Page 98-101; 144pp; English.
XX
CC   The present invention relates to Corynebacteium glutamicum
CC   phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC   The PTS nucleic acids and proteins are useful in the
CC   identification of microorganisms which can be used to produce fine
CC   chemicals, for modulating fine chemical production in C. glutamicum or
CC   related bacteria, the typing or identification of C. glutamicum or
CC   related bacteria, as reference points for mapping C. glutamicum genome,
CC   and as markers for transformation.
XX
SQ   Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;


  Query Match             100.0%;  Score 1527;  DB 22;  Length 1527;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1527;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500

Qy   1501 accataagttgaaaccttgagtgttcg 1527
          |||||||||||||||||||||||||||
Db   1501 accataagttgaaaccttgagtgttcg 1527


RESULT   2
AAH68533/c
ID   AAH68533 standard; DNA; 349980 BP.
XX
AC   AAH68533;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 7068.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
```

PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Disclosure; SEQ ID NO: 7068; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;

Query Match 99.2%; Score 1515; DB 22; Length 349980;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1526; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111326 CTCATGGCATCTGCGCCGTTCGCGTTCTTGCCAGTGTTGGTTGGTTTCACCGCAACCAAG 111267

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 111266 CGTTTCGGCGGCAATGAGTTCCTGGGCGCCG-GTATTGGTATGGCGATGGTGTTCCCGAG 111208

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111207 CTTGGTGAACGGCTACGACGTGGCCGCCACCATGGCTGCGGGCGAAATGCCAATGTGGTC 111148

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111147 CCTGTTTGGTTTAGATGTTGCCCAAGCCGGTTACCAGGGCACCGTGCTTCCTGTGCTGGT 111088

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111087 GGTTTCTTGGATTCTGGCAACGATCGAGAAGTTCCTGCACAAGCGACTCAAGGGCACTGC 111028

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111027 AGACTTCCTGATCACTCCAGTGCTGACGTTGCTGCTCACCGGATTCCTTACATTCATCGC 110968

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110967 CATTGGCCCAGCAATGCGCTGGGTGGGCGATGTGCTGGCACACGGTCTACAGGGACTTTA 110908

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110907 TGATTTCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGGTCTACTCACCAATCGTCAT 110848

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110847 CACTGGTCTGCACCAGTCCTTCCCGCCAATTGAGCTGGAGCTGTTTAACCAGGGTGGATC 110788

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110787 CTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCCAGGGTGCGGCATGTTTGGCAGT 110728

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110727 GTTCTTCCTGGCGAAGAGTGAAAAGCTCAAGGGCCTTGCAGGTGCTTCAGGTGTCTCCGC 110668

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110667 TGTTCTTGGTATTACGGAGCCTGCGATCTTCGGTGTGAACCTTCGCCTGCGCTGGCCGTT 110608

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110607 CTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATCAA 110548

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110547 GGCAGTTGCGTTGGGCGCTGCAGGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGATAT 110488

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110487 GGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATTGC 110428

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110427 TTATGGCCTTTACTTGGTTCGCCGCAACGGCAGCATTGATCCAGATGCAACCGCTGCTCC 110368

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110367 AGTGCCTGCAGGAACGACCAAAGCCGAAGCAGAAGCACCCGCAGAATTTTCAAACGATTC 110308

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110307 CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGATGC 110248

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110247 CATGTTTGCCAGCGGAAAGCTTGGCTCGGGCGTTGCCATCGTCCCAACCAAGGGGCAGTT 110188

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110187 AGTTTCTCCGGTGAGTGGAAAGATTGTGGTGGCATTCCCATCTGGCCATGCTTTCGCAGT 110128

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110127 TCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGCACATTGGTTTCGACAC 110068

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110067 AGTAAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAAGC 110008

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110007 AGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAGGCTGCAGGTTATGAGGTAAC 109948

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109947 CACGCCGATTGTTGTTTCGAATTACAAGAAAACCGGACCTGTAAACACTTACGGTTTGGG 109888

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109887 CGAAATTGAAGCGGGAGCCAACCTGCTCAACGTCGCAAAGAAAGAAGCGGTGCCAGCAAC 109828

Qy   1501 accataagttgaaaccttgagtgttcg 1527
          |||||||||||||||||||||||||||
Db 109827 ACCATAAGTTGAAACCTTGAGTGTTCG 109801
```

```
RESULT   3
AAH67869
ID   AAH67869 standard; DNA; 1983 BP.
XX
AC   AAH67869;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 2904.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG92650.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Claim 8; SEQ ID NO: 2904; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 1983 BP; 410 A; 508 C; 558 G; 507 T; 0 other;

  Query Match            97.7%;  Score 1492;  DB 22;  Length 1983;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1503;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 540

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    541 cgtttcggcggcaatgagttcctgggcgccg-gtattggtatggcgatggtgttcccgag 599

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 659

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 719

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 779

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 839

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 899

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 959

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 1019

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 1079

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 1139

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 1199

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 1259

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 1319

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 1379

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 1439

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1499

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1559

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1619

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1679

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1739

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1799

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1859

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1919

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1979

Qy   1501 acca 1504
          ||||
Db   1980 acca 1983


RESULT   4
AAF32543
ID   AAF32543 standard; DNA; 5969 BP.
XX
AC   AAF32543;
XX
DT   20-APR-2001  (first entry)
XX
DE   Brevibacterium lactofermentum sucrose PTS enzyme II DNA SEQ ID NO:1.
XX
KW   Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrase;
KW   phosphoenolpyruvate:carbohydrate phosphotransferase system; glucose;
KW   coryneform bacterium; phosphoenolpyruvate-sugar transport system; ds.
XX
OS   Brevibacterium lactofermentum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3779..5764
FT                   /*tag= a
FT                   /product= "sucrose PTS enzyme II"
XX
PN   WO200102584-A1.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-JP04348.
XX
PR   02-JUL-1999;   99JP-0189512.
XX
PA   (AJIN ) AJINOMOTO CO INC.
XX
PI   Izui M,  Sugimoto M,  Nakamatsu T,  Kurahashi O;
XX
DR   WPI; 2001-138150/14.
DR   P-PSDB; AAB69080.
XX
PT   Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT   II obtained by cassette ligation-mediated amplification of downstream
PT   domain of coryneform bacterium sucrase gene, with sucrose-binding
PT   activity -
XX
PS   Claim 3; Page 22-29; 45pp; Japanese.
XX
CC   The present sequence encodes the Brevibacterium lactofermentum sucrose
CC   PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system or
CC   phosphoenolpyruvate-sugar transport system) enzyme II, which has
CC   sucrose-binding activity. A coryneform bacteria produced with the
CC   sucrose PTS enzyme II gene can have more efficient sugar uptake, and
CC   improved amino-acid and nucleic acid productivity. The sucrose PTS gene
CC   and it's disrupted gene, such as one without the sucrose PTS function,
CC   can be used to produce new breeds of coryneform bacterial strains to
CC   uptake sugar more efficiently e.g. glucose only or and sucrose, and can
CC   have improved amino-acid and nucleic acid productivity.
XX
SQ   Sequence 5969 BP; 1388 A; 1575 C; 1578 G; 1428 T; 0 other;


  Query Match            92.5%;  Score 1412.6;  DB 22;  Length 5969;
  Best Local Similarity  95.7%;  Pred. No. 0;
  Matches 1462;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4259 ctgatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 4318

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          |||||||| |||||||||||||||||||||| | ||||||||||||||||||||||| |
Db   4319 cgtttcggtggcaatgagttcctgggcgccg-gcattggtatggcgatggtgttcccaac 4377

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          | |||| |||||||||||||||||||||||| | ||||||||||||||||||||||
Db   4378 cctggttaacggctacgacgtggccgccaccatgaccgcgggcgaaatgccaatgtggtc 4437

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          |||||||||||| |||||||| ||||| ||||||||||||||||||||||||||||||||
Db   4438 cctgtttggtttggatgttgctcaagctggttaccagggcaccgtgcttcctgtgctggt 4497

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   4498 ggtctcttggattctggcaacgatcgagaagttcctgcacaagcgactcatgggcactgc 4557

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          |||||||||||||||| |||||| |||| |||||||||||||| |||||||| ||||| ||
Db   4558 agacttcctgatcaccccagtgttgactctgctgctcaccggcttccttacgttcattgc 4617

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||| |||||||||||||||||||| ||  |||||||||||||||| |||||||| ||
Db   4618 tattggtccagcaatgcgctgggtgggtgacttgctggcacacggtctgcagggactcta 4677

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db   4678 tgatttcggtggtccagtcggcggtctgcttttcggtctggtctactcaccaatcgttat 4737

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4738 cactggtctgcaccagtccttcccgccaattgagctggagctgttcaaccagggtggatc 4797

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          |||||||||||||| ||||| ||||| |||||||| |||||||| ||||||||||||||||
Db   4798 cttcatcttcgcaaccgcatccatggccaatatcgcgcagggtgcagcatgtttggcagt 4857

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4858 gttcttcctagcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 4917

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   4918 tgttcttggtattacagagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 4977

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          || ||| |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   4978 ctacattggtatcggtaccgcagctatcggtggcgctttgattgcactctttgatatcaa 5037

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5038 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 5097

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          |||||||||||||| || || || |||||||| ||||||||||||||||||||||||||
Db   5098 ggtcatgttcttggtttgcgcggtagttacctttgtcatcgcattcggcgcagcgattgc 5157

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5158 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 5217

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
```

```
Db   5218 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 5277

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
Db   5278 caccatcatccaggcacctttgaccggtgaagctatcgcactgagcagcgtcagcgatgc 5337

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
Db   5338 catgtttgccagcggaaagcttggctcaggtgttgcgatcgtccccaccaaggggcagct 5397

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
Db   5398 ggtttcaccagtgagcggaaagatcgtggtggccttcccatctggtcacgctttcgcagt 5457

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
Db   5458 ccgcactaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 5517

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
Db   5518 cgtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 5577

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
Db   5578 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 5637

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
Db   5638 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 5697

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
Db   5698 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 5757

Qy   1501 accataagttgaaaccttgagtgttcg 1527
Db   5758 accataagttgaaaccttgagtgttcg 5784
```

RESULT 5
AAF31529
ID AAF31529 standard; DNA; 1109 BP.
XX
AC AAF31529;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate DNA #2.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-080989/09.
XX
PT Corynebacteium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation -
XX
PS Claim 5; Page 103-104; 144pp; English.
XX
CC The present invention relates to Corynebacteium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.
XX
SQ Sequence 1109 BP; 236 A; 278 C; 303 G; 292 T; 0 other;

Query Match 72.5%; Score 1107.4; DB 22; Length 1109;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    419 tatgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtc 478
Db      1 tatgatttcggcggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtc 60

Qy    479 atcactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtgga 538
Db     61 atcactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtgga 120

Qy    539 tccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggca 598
Db    121 tccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggca 180

Qy    599 gtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctcc 658
Db    181 gtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctcc 240

Qy    659 gctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccg 718
Db    241 gctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccg 300

Qy    719 ttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatc 778
Db    301 ttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatc 360

Qy    779 aaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagat 838
Db    361 aaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagat 420

Qy    839 atggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgatt 898
Db    421 atggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgatt 480

Qy    899 gcttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgct 958
Db    481 gcttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgct 540

Qy    959 ccagtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgat 1018
Db    541 ccagtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgat 600

Qy   1019 tccaccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgat 1078
Db    601 tccaccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgat 660

Qy   1079 gccatgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcag 1138
Db    661 gccatgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcag 720

Qy   1139 ttagtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgca 1198
Db    721 ttagtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgca 780
```

```
Qy   1199 gttcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgac 1258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 gttcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgac 840

Qy   1259 acagtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaa 1318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 acagtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaa 900

Qy   1319 gcaggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggta 1378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 gcaggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggta 960

Qy   1379 accacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttg 1438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 accacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttg 1020

Qy   1439 ggcgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagca 1498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ggcgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagca 1080

Qy   1499 acaccataagttgaaaccttgagtgttcg 1527
          |||||||||||||||||||||||||||||
Db   1081 acaccataagttgaaaccttgagtgttcg 1109
```

```
RESULT   6
AAH54100/c
ID   AAH54100 standard; DNA; 2913 BP.
XX
AC   AAH54100;
XX
DT   03-SEP-2001  (first entry)
XX
DE   S. epidermidis genomic polynucleotide sequence SEQ ID NO:3464.
XX
KW   Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW   vaccination; endocarditis; ds.
XX
OS   Staphylococcus epidermidis.
XX
PN   WO200134809-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30782.
XX
PR   09-NOV-1999;   99US-0164258.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Kimmerly WJ;
XX
DR   WPI; 2001-316495/33.
XX
PT   Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT   useful for vaccinating against infections, e.g. endocarditis -
XX
PS   Claim 8; Page 1030-1031; 2188pp; English.
XX
CC   AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC   (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC   (I) and (II) can have antibacterial activity and therefore can be used
CC   in vaccination. The nucleic acids (I) may be used to produce the
CC   S. epidermidis polypeptides (II) via the production of vectors
CC   containing them which are used to produce hosts cells which express the
CC   polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC   used to vaccinate subjects and to raise antibodies against the bacteria.
CC   The polypeptides may also be used to assay for other inhibitors of their
CC   activity and therefore identify compounds that may be used for the
CC   treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC   AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC   polynucleotide sequences from the present invention. AAH55091 to
CC   AAH55098 represent oligonucleotide sequences and primers which are used
CC   in the exemplification of the present invention.
CC   N.B. The present invention specifically claims all the polynucleotide
CC   sequences given in the sequence listing of the present specification,
CC   however the sequence listing only goes up to SEQ ID NO:4454 so even
CC   though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC   no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ   Sequence 2913 BP; 1072 A; 480 C; 419 G; 942 T; 0 other;
```

```
  Query Match             12.5%;  Score 190.6;  DB 22;  Length 2913;
  Best Local Similarity   55.0%;  Pred. No. 6.2e-46;
  Matches  468;  Conservative    0;  Mismatches  364;  Indels   19;  Gaps    4;

Qy       5 tggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcgtt 64
           | |||  ||| || ||  | || || ||| |  |  |||||||  |  ||| | ||||| |
Db    2690 TTGCAAATGCACCTTTTACATTATTACCAATACTTATTGGTTTTAGTGCAGCTAAGCGAT 2631

Qy      65 tcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagcttg 124
           | || || |||   |   | || || || | | ||||||   | ||   || | |||
Db    2630 TTGGTGGTAATCCTTATTTAGGTGCAGC-TCTAGGTATGATACTTGTTCATCCTGGATTG 2572

Qy     125 gtgaacggctacgacgtggccgccaccatggctgcgggcgaa---atgccaatgtggtcc 181
           |||  |  || ||  |  |     |  | |  |  ||  ||   || |||   |||
Db    2571 ATGAGTGCATATGATTTTCCAAAAGCTTTAGAAGAAGGAAAAGCTATTCCACACTGGGAT 2512

Qy     182 ctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtg 241
           | |||||  || || ||   |||  ||||| |||||    ||  | ||| |||| ||
Db    2511 GTCTTTGGGCTACATATTAATGAAGTAGGTTATCAGGGACAAGTATTACCTATGCTCGTA 2452

Qy     242 gtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgca 301
           |   | |  ||| | || ||||| || || |   | |  ||     |           |
Db    2451 GCAACATATATTTTAGCTACGATTGAAAAATGGTTACGTAAAGTTATTCCAACTGTGTTA 2392

Qy     302 gacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcc 361
           ||     |  | || ||| |  |  |  |  |  | || | |||  | |||||  |
Db    2391 GATAATTTATTGACGCCATTATTATCAATTTTTATTACAGCATTTATAACATTTTTATTT 2332

Qy     362 attggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttat 421
           | || || |  |  ||     | ||  ||  | |  |  | ||  |     |  | |||
Db    2331 GTAGGGCCTGTCACTCGTCAATTAGGTTATTGGTTATCTGATGGATTGACTTGGTTATAT 2272

Qy     422 gatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatc 481
           || || |||||  |  | || || ||  | || ||| |  | ||  | ||||||||||||
Db    2271 GAATTTGGTGGAGCTATTGGTGGACTTATATTTGGTTTATTATATGCGCCAATCGTCATT 2212

Qy     482 actggtctgcaccagtccttcccgccaattgagctg-----------gagctgtttaac 529
           || ||  |||||||   |||     || ||||  |             ||  |  |||
Db    2211 ACAGGAATGCACCATAGCTTTATTGCAGTTGAAACGACATTAATTGCTGATGCGACTAAA 2152

Qy     530 cagggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggca 589
             ||||| || || |||||| |||  ||| | ||| | ||||| || || ||||  |||
Db    2151 ACAGGTGGTTCATTTATCTTCCCAATCGCAACGATGTCAAATATTGCACAAGGTGGTGCA 2092

Qy     590 tgtttggcagtgttct---tcctggcgaagagtgaaaagctcaagggccttgcaggtgct 646
           |||| || |  ||||   || |  | | | | ||||  | || ||  ||||   ||
Db    2091 GCTTTAGCTGCATTCTTTATCATTAAGCAAAATAAAAAATTAAAAGGTGTTGCTTCCGCG 2032

Qy     647 tcaggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgc 706
           | ||| | || ||| | || || ||||| || || || || || ||||| || |||
Db    2031 GCGGGTATTTCAGCTTTACTAGGAATTACAGAACCAGCAATGTTTGGTGTCAATCTTAAA 1972

Qy     707 ctgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgca 766
           || | |  || ||  |    | | | |       ||| ||| ||||  ||||  || ||
Db    1971 TTGAGATATCCATTTATAGGTGCTGTTGCAGGATCAGGTATAGGTGCGGCTTATATTTCA 1912

Qy     767 ctctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctatt 826
           ||||  ||  | || ||  | ||| | ||  | || || ||    ||  || | |||||
Db    1911 TTCTTCAAAGTAAAAGCGATAGCGCTTGGTACAGCTGGATTACCTGGATTTATATCTATA 1852

Qy     827 gatgctccaga 837
           || || || |
Db    1851 AATCCTACACA 1841
```

```
RESULT   7
AAX12968
ID   AAX12968 standard; DNA; 5840 BP.
XX
AC   AAX12968;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:31.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine;,attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 389-392; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 5840 BP; 1848 A; 944 C; 1256 G; 1779 T; 13 other;

  Query Match             9.9%;  Score 151.2;  DB 20;  Length 5840;
  Best Local Similarity   48.5%;  Pred. No. 4.9e-34;
  Matches  525;  Conservative    0;  Mismatches  533;  Indels   24;  Gaps    3;

Qy     329 ttgctgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggtgggc 388
           || ||  | |  | |    |||| ||| | || ||||| ||       |      | |||
Db    1850 ttccttattatggcacctgttaccttcttagctattggtcctatcgggacggtcatcggc 1909

Qy     389 gatgtgctggcacacggtctacagggactttatgatttcggtggtccagtcggcggtctg 448
           |||  | ||| ||| ||     | | | |||| | |||  ||       |||  ||  |
Db    1910 gattggttgggacaaggatacaacgcaatttacgcttttagtccaattatcgcagggtta 1969

Qy     449 ctcttcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttcccgcca 508
            |  | |||  | | |    |    | || ||   |||| ||||   |  |||    |||
Db    1970 ttaatgggttcgttgtggcaagtcttggtaatgtttggtatgcattggggctttgtacca 2029

Qy     509 attgagctggagctgtttaaccagggtggatccttcatcttcgcaacggcatctatggct 568
           |||  | ||      || |  || |||||       ||  |  | | |  ||     ||
Db    2030 attatgatgttaaacttaacacaaggtggcgatacgatggtaccgatgttattaccagcc 2089

Qy     569 aatatcgcccagggtgcggcatgtttggcagtgttcttcctggcgaagagtgaaaagctc 628
             ||| || || || | |||   ||| || || || |||||  | || | ||  || ||
Db    2090 gttattgcacaaggcggggctgctttagctgtctttttcctaacaaaaaatgtgaaacta 2149

Qy     629 aagggccttgcaggtgcttcaggtgtctccgctgttcttggtattacggagcctgcgatc 688
           || ||  | ||     ||||| || |  |  || || |||| ||||| || ||  |  |
Db    2150 aaaggtttggctttgtcttcaagtattacgactatttttggaattactgaaccaactgta 2209

Qy     689 ttcggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagctatc 748
           |  || ||||   | |   ||      || ||  |    | |     |   |  | ||||
Db    2210 tatggcgtgactttaccattgaaaaaaccatttattgcagcttgtattggtggcggtatc 2269

Qy     749 ggtggcgctttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggtttc 808
           ||||| || ||  | ||  |   | |  | ||     || |||| |||   |   || |
Db    2270 ggtggtgcatttgtggctatgaatcacgtgaaaaactttacgtttggcttggttagtatg 2329

Qy     809 ttgggtgttgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagttgtt 868
           ||| |  |   |    ||  | || |||| |                |        || |
Db    2330 ttgagcttgcctggctttattcctgcagagacaaaagatactgcaccgatgattactggt 2389

Qy     869 accttcttcatcgcattcggcgcagcgattgcttatggcctttacttggttcgccgcaac 928
            |  |      || | | | |      ||||| | || | |    || ||   ||
Db    2390 gcaattggtgccggaattgcctttatcattgcgtttgtcttaacgtttgtcttacgtt-- 2447

Qy     929 ggcagcattgatccagatgcaaccgctgctccagtgcctgcaggaacgaccaaagccgaa 988
                ||||     |  | |   | |   |||   ||| |  |||    |  |   |
Db    2448 -------ttgaagatcaacctaatccagaaacagcaactgaaaaaacagaaactgataag 2500

Qy     989 gcagaagcacccgcagaattttcaaacgattccaccatcatccaggca---cctttgacc 1045
           |  ||||| | | ||       | ||   ||  || ||    |||   ||  |
Db    2501 atggtggcacctgtaaaaacgaatcaagaagacaaaattattttagcaagtccacttcaa 2560

Qy    1046 ggtgaagctattgcactgagcagcgtcagcgatgccatgtttgccagcggaaagcttggc 1105
           ||||||  | |  | ||    |  ||    ||  |  | |||||    ||     | ||
Db    2561 ggtgaaatttaccgctagaaaaagtacaagaccctgtttttgcttcaggtgctttagga 2620

Qy    1106 tcgggcgttgccatcgtcccaaccaaggggcagttagtttctccggtgagtggaaagatt 1165
             || ||||| || |  || ||  | ||  |  |   | | || |    |||  | ||
Db    2621 aaaggtgttgcaattgagccgactgaaggcaaactgtatgcacccgcagatggtgaaatc 2680

Qy    1166 gtggtggcattcccatctggccatgctttcgcagttcgcaccaaggctgaggatggttcc 1225
                ||| ||  | || |||||| | |    |     || |  |  |
Db    2681 accacattatttccgacaggacatgctgttggcttgacgacaacagagg----------- 2729

Qy    1226 aatgtggatatcttgatgcacattggtttcgacacagtaaacctcaacggcacgcacttt 1285
             || ||  | || ||||| |||||  | || || ||  |  |  | || |    ||||
Db    2730 -gcgttgaattattaatgcatattggcatggatacggtcgaattagatggtaaaggcttt 2788

Qy    1286 aacccgctgaagaagcagggcgatgaagtcaaagcaggggagctgctgtgtgaattcgat 1345
            |       ||| || || |||   || |||  ||| ||  ||||   |   || |||
Db    2789 gaattatcagtgaaacaaggtgattctgttaaaaaaggagatttgctagttacttttgat 2848

Qy    1346 attgatgccattaaggctgcaggttatgaggtaaccacgccgattgttgtttcgaattac 1405
           |||| |||||||| |  || ||||||  ||||   || |||||||| ||  |||||
Db    2849 attgctgccattaaagaagctggttatccggtagttacaccgattgtggtaacgaatacg 2908

Qy    1406 aa 1407
           ||
Db    2909 aa 2910


RESULT   8
ABA03041
ID   ABA03041 standard; DNA; 2944528 BP.
XX
AC   ABA03041;
XX
DT   05-FEB-2002  (first entry)
```

XX
DE Listeria monocytogenes EGD-e genome sequence.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease; ds.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 1; SEQ ID No 1; 192pp; French.
XX
CC The present sequence is the genome sequence of Listeria monocytogenes
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in L. monocytogenes and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Proteins (ABB47297-ABB50149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of L.
CC monocytogenes and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate L.
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccines compositions for
CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match 8.7%; Score 132.2; DB 24; Length 2944528;
Best Local Similarity 47.6%; Pred. No. 6.4e-27;
Matches 558; Conservative 0; Mismatches 573; Indels 42; Gaps 4;

```
Qy    316 tccagtgctgacgttgctgctcaccggattccttacattcatcgccattggcccagcaat 375
          |||| |    || ||  |  ||   |      | || |||||||  |||||||||
Db  29873 tccattctgtacattattaatcgtagttccaatcactttcatcgtaattggcccaattgc 29932

Qy    376 gcgctgggtgggcgatgtgctggcacacggtctacagggactttatgatttcggtggtcc 435
            ||||  ||  |  |  | | |   ||   |    |  ||||| | ||  ||
Db  29933 tacatgggctggtcaattattaggagcaggaacaatttgggtttataacttaagtccaat 29992

Qy    436 agtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcactggtctgcacca 495
           |||  |||||  | || |||    |||    |  | | || |||  ||| || |||
Db  29993 tatcgcaggtctaattttaggtggtttctggcaagtatttgttatctttggacttcactg 30052

Qy    496 gtccttcccgccaattgagctggagctgtttaaccagggtggatccttc---atcttcgc 552
                |    |   ||| | |   |  |    || |       ||||         || || ||
Db  30053 gggtcttgtaccagtggcaatcaacaatttaactgttcttggacatgatccaatttttagc 30112

Qy    553 aacggcatctatggctaatatcgcccagggtgcggcatgtttggcagtgttcttcctggc 612
          || | ||| |   ||   | | || ||    |  |||   |||||||| |||||     |
Db  30113 aatgacatttggtgcatcttttgctcaaatcggtgcagtattggcagtattctttaaatc 30172

Qy    613 gaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttggtat 672
          |  |   ||||  ||||    |||       |  ||  | | |||| | |  ||||| |
Db  30173 tagaaacaaaaaaatcaaatctcttagtatcccagcatttatttccggtatctttggtgt 30232

Qy    673 tacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcggtat 732
           || || ||  |||||  | ||||||  |   | |  |||      || ||  ||||  |
Db  30233 aactgaaccagcgatttacggtgttactttacctctgaaaaaaccatttatcatgagctg 30292

Qy    733 cggtaccgcagctatcggtggcgctttgattgcactctttaatatcaaggcagttgcgtt 792
              | |  | |||  |||||||   | ||||   |           |  ||  |    |
Db  30293 tatcgctggtggtattggtggcggaatcattggttttgcaggatctcaaacatatatcat 30352

Qy    793 gggcgctgcaggtttcttgggtgttgtttctattgatgctccagatatggtcatgttctt 852
          ||| |     ||  |||| ||| ||  |    |       |||| ||  |  || |
Db  30353 gggtggacttggaatcttcggtcttcctaacttcttcaaaccaggtagcggtatttcagg 30412

Qy    853 ggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgcttatggccttta 912
          | |          ||    ||  ||||  || ||       ||| ||    | |
Db  30413 cgaattctggtgggtagttatcgcaatcgttatcagcttcatccttggtttcatcttaac 30472

Qy    913 cttggttcgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctgcagg 972
          |  ||                | | |    | |   | |||| |  |||    | |
Db  30473 ttatgt---------------agtaggtttcaaagatccagctgatgtagtagtagaaca 30517

Qy    973 aacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattccaccatcatcca 1032
          | | | || ||  ||||  ||| |           || || |||       |  || |
Db  30518 atctaacacagtagaaggcgaaac------------tttaatcgaacgtgaaacaattcc 30565

Qy   1033 ggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttgccag 1092
          || ||  |    || |||     |  |  |       || |  || |     ||| |
Db  30566 agctccagtagttggcgaaatcgtaactttagcagatgtaaaagacgaagcattttcatc 30625

Qy   1093 cggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctccggt 1152
          |||    |||||    || |||| ||  |||| ||    || |   | ||  |||| |
Db  30626 cggtgcacttggaaaaggtgttgcaattatccctactgtcggacgcgttgtagctccagc 30685

Qy   1153 gagtggaaagattgtggtggcattcccatctggccatgctttcgcagttcgcaccaaggc 1212
            ||||  |          |||||| |||| |||||  |||   ||   || || |
Db  30686 agccggaacagtaacaactatcttcccaactggtcatgcaatcggtattacaacaaaaga 30745

Qy   1213 tgaggatggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacctcaa 1272
          ||              ||  ||  | || || ||||| ||| | ||||| ||  | || |
Db  30746 tgg-----------tgcagaagttttaatccacatcggtatggacactgttcaacttga 30793

Qy   1273 cggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagctgct 1332
          || |    ||| |   | |     ||||| ||||| | | || ||  |||  |  | ||
Db  30794 aggtaaattcttcacagcacacgttaagcaaggcgacgtaatcgaaaaaggtcaattact 30853

Qy   1333 gtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaaccacgccgattgt 1392
            |||||| |||||||| | |||||| || || || ||||| || || || ||  | ||
Db  30854 aactgaatttgatattgaaggcattaaagcagctggatatgacgttacaactccagtagt 30913

Qy   1393 tgtttcgaattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattgaagc 1452
          ||  | || | |||  ||    |  ||||  | |   | |       |||    ||
Db  30914 agtaacaaactccaatcaatatttagacgtaatgattacagatgcaaaagaagcaaaact 30973

Qy   1453 gggagccaacctgctcaacgtcgcaaagaaaga 1485
          | ||      |  |||   | | ||   ||||
Db  30974 agaagaacgtttaatcacactagtaatctaaga 31006
```

RESULT 9
AAV52334/c
ID AAV52334 standard; DNA; 3895 BP.
XX
AC AAV52334;

XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:201.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
PS Claim 1; Page 1179-1182; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 3895 BP; 1166 A; 869 C; 805 G; 1055 T; 0 other;

Query Match 8.3%; Score 126.6; DB 19; Length 3895;
Best Local Similarity 47.1%; Pred. No. 8.4e-27;
Matches 575; Conservative 0; Mismatches 619; Indels 27; Gaps 5;

```
Qy    190 tttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtggtttcttg 249
          |||   ||||   |   |||||||||| | |   ||  | || |   || | |    ||
Db   3218 TTTTACTGTTAATCGTATCGGTTACCAAGCCCAAGTTATCCCAGCCTTGCTTGCAGGTTT 3159

Qy    250 gattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcagacttcct 309
          || |||| |    | || |  |||  |||||||||   ||   |          |  |
Db   3158 GAGTCTGTCTTATCTTGAAATCTTCTGGCACAAGCATATCCCAGAAGTCATTTCTATGAT 3099

Qy    310 gatcactccagtgctgacgttgctgctcaccggattccttacattcatcgccattggccc 369
            |    ||  |  || | ||| | |   ||    |  |  |    |  |   | || ||
Db   3098 TTTTGTACCTTTCTTGTCATTGATTCCAGCCTTGATTTTGGCTCATACTGTTTTGGGACC 3039

Qy    370 agcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttt---atgattt 426
          |       |     | ||  | |  ||  ||   | | |   || |  ||    || |
Db   3038 AATCGGTTGGACAATTGGACAAGGACTTTCATCAGTTGTCTTGGCAGGTTTAACTGGTCC 2979

Qy    427 cggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcactgg 486
          | |       || ||||    | || ||     ||||| | ||| | |||||||| ||
Db   2978 AGTTAAATGGCTCTTCGGTGCAATTTTTGGCGCCCTCTACGCTCCATTTGTCATCACAGG 2919

Qy    487 tctgcaccagtccttcccgccaattgagctggagctgttt---aaccagggtggatcctt 543
          |||||||||      |    | |||||     |  || ||    |    |||||  |
Db   2918 TCTGCACCATATGACCAATGCCATTGATACACAATTGATTGCGGATGCTGGTGGCACTGC 2859

Qy    544 catcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagtgtt 603
          | | |   ||| |    || |  ||||||| || || ||  | ||    || ||    |
Db   2858 CCTATGGCCAATGATTGCTCTTTCTAATATTGCTCAAGGCTCAGCCGTGTTTGCCTATTA 2799

Qy    604 cttcctggcgaagagtgaaaagctcaagggcc---ttgcaggtgcttcaggtgtctccgc 660
          |||  ||       |||  |||   |||  |   || ||  | || ||    | || ||
Db   2798 TTTCATGCATCGCCATGATGAGCGTGAGGCTCAGGTTTCACTTCCTGCAACCATTTCAGC 2739

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
            ||| ||| |||| || || ||  | || || || ||| |           |  || ||
Db   2738 CTATCTCGGTGTTACAGAACCAGCTCTTTTTGGGGTTAACGTAAAATATATTTATCCATT 2679

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          |    || ||   |    ||||  | |  |||   ||    |     ||||| | |
Db   2678 TGTTGCTGGGATGACTGGTTCAGCCCTTGCAGGCATGTTATCCGTTACTTTTAATGTAAC 2619

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          || | | |  | ||    |  |||||    ||| || | |||||| |  ||| | | ||
Db   2618 TGCGGCTTCTATTGGTATCGGTGGTTTGCCAGGTATTCTCTCTATTCAACCTCAATACAT 2559

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          | |   ||    |    |   | ||| |  | |  |  |  || |   |      | |
Db   2558 GCTGCCATTTGCAGGAACTATGCTAGTTGCGATTGTTGTTCCAATGCTC------TTGAC 2505

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||    | |   |  || | | | |       | |  ||    ||| | | |       |
Db   2504 TTTCTTCTTCCGCAAGGCTGGTCTCTTTACAAAAACAGAGGGCGATACGAACTTGCAGGC 2445

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
          ||    |  |      |||| |||    |    | | | | |  |     |||
Db   2444 AGAATTCGTTGCTCAAGAAGAAGCAGAATTTGTGAACCATGAACCAGTAGAACTTACTTC 2385

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
               ||       ||  | || ||  |||  |  | |  ||||    |  |  ||| |
Db   2384 GGTAGAAATTATCAGCCCACTAACTGGCCAAGTGAAAGAATTGAGTCAAGCGACGGATCC 2325

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
           || |||||    ||      | ||    ||  | | ||| |  |||| | | ||  ||||
Db   2324 TATTTTTGCATCAGGTGTCATGGGGCAAGGTCTAGTCATTGAACCAAGCCAAGGTGAGTT 2265

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
               ||||| || | ||| |   |   |||   |||||  |   |||||  |||   |
Db   2264 GACCTCTCCAGTTAATGGGACAGTGACGGTTCTTTTCCCTACCAAGCATGCCATCGGCAT 2205

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          |  | |  | |  |                 || ||  |  | || ||||| ||| | || ||
Db   2204 TGTCTCTGACGAGG-----------GAGTTGAATTGCTCATCCACATCGGTATGGATAC 2157

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
          ||||   ||  | ||||     ||| |    ||       || || ||| | || | ||
Db   2156 AGTAGGTCTTGATGGCAAAGGTTTTGAAAGTCTTGTAGTCCAAGGAGATCACGTTACAGT 2097

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
           ||  |||  |||  |   || ||||| |||| ||||||||||||||||   | |  |
Db   2096 TGGTCAGCAACTGATTCGTTTTGATATGGATGTCATTAAGGCTGCAGGTCTGGTGACAGA 2037

Qy   1381 cacgccgattgttgtttcgaa 1401
           || ||  || |  |  | ||
Db   2036 AACTCCTGTTATCATCACCAA 2016
```

```
RESULT  10
AAS55732
ID   AAS55732 standard; DNA; 1884 BP.
XX
AC   AAS55732;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae DNA for cellular proliferation protein #303.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU37873.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 9369; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1884 BP; 478 A; 395 C; 435 G; 576 T; 0 other;


  Query Match             7.4%;  Score 113.4;  DB 23;  Length 1884;
  Best Local Similarity   46.8%;  Pred. No. 4.8e-23;
  Matches  560;  Conservative    0;  Mismatches  601;  Indels   36;  Gaps    5;

Qy    209 ggttaccagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgatcgag 268
Db    640 ggtttgcaaggttccgttcttccagccttcatcatcggggttgtcggagctaaatttgaa 699

Qy    269 aagttcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgctgacg 328
Db    700 aaagctgtccgcaaggttgttccagatgtcattgacctcttggtaacaccattcgtgaca 759

Qy    329 ttgctgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggtgggc 388
Db    760 cttttggtcatgtctatccttggactctttgtcattggaccagttttccacgttgttgaa 819

Qy    389 gatgtgctggcacacggtctacagggactttatgatttcggtggtccagtcggcggtctg 448
Db    820 aactacatccttattgctacaaaagcgattcttagcatgccatttggtcttggtggtttc 879

Qy    449 ctcttcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttcccgcca 508
Db    880 ttgattggtggggttcaccaattgatcgtcgtgtcaggtgtgcaccacatcttcaacttg 939

Qy    509 attgagctggagctgtttaaccagg---gtggatccttcatcttcgcaacggcatctatg 565
Db    940 cttgaagtgcaattacttgctgctgaccatgctaacccattcaacgctatcatcacagct 999

Qy    566 gctaatatcgcccagggtgcggcatgtttggcagtgttcttcctggcgaagagtgaaaag 625
Db   1000 gctatgacagctcaaggtgctgctactgttgcggttggtgttaaaacaaaaaatccaaaa 1059

Qy    626 ctcaagggccttgcaggtgcttcaggtgtctccgctgttcttggtattacggagcctgcg 685
Db   1060 ctgaaaacacttgctttcccggctgctctttctgccttcctaggtattacagagcctgct 1119

Qy    686 atcttcggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgcagct 745
Db   1120 atcttcggggtgaacttgcgcttccgtaaaccattcttcctttcattgattgctggtgca 1179

Qy    746 atcggtggcgctttgattgcactctttaatatcaaggcagttgcgttgggcgctgcaggt 805
Db   1180 atcggtggtggattggcttctatccttggacttgctggtactggtaatggtatcaccatc 1239

Qy    806 ttcttgggtgttgtttctattgatgctccagatatggtcatgttcttggtgtgtgcagtt 865
Db   1240 atccctggtacaatgctttatgttggtaacggacaacttccacaataccttcttatggta 1299

Qy    866 gttaccttcttcatcgcattcggcgcagcgattgcttatggcctttacttggttcgccgc 925
Db   1300 gctgtatcatt----------------tgccctggttttgctcttacttacatgtttggt 1344

Qy    926 aacggcagcattgatccagatgcaaccgctgctccagtgcctgcaggaacgaccaaagcc 985
Db   1345 tacg---aagatgaagtagacgcaactgcagctgcaaaacaagctgaagtggctgaagaa 1401

Qy    986 gaagcagaagcacccgcagaatttttcaaacgattccaccatcatccaggcacctttgacc 1045
Db   1402 aaagaagaagttgcgccagcagctcttcaaaatgaaacacttgt---aactcctatcgtc 1458

Qy   1046 ggtgaagctattgcactgagcagcgtcagcgatgccatgtttgccagcggaaagcttggc 1105
Db   1459 ggtgatgttgtcgctcttgctgatgtcaatgacccagtcttctcaagtggagctatggga 1518

Qy   1106 tcgggcgttgccatcgtcccaaccaaggggcagttagtttctccggtgagtggaaagatt 1165
Db   1519 caaggtatcgctgtgaaaccaagtcaaggcgtggtctatgcactagctgatgctgaagtt 1578

Qy   1166 gtggtggcattcccatctggccatgctttcgcagttcgcaccaaggctgaggatggttcc 1225
Db   1579 tcaattgcctttccaacagggcacgcttttggtttgaaaacaagaaatg---------- 1627

Qy   1226 aatgtggatatcttgatgcacattggtttcgacacagtaaacctcaacggcacgcacttt 1285
Db   1628 -gtgctgaagttttgattcacgttggtattgacactgtatctatgaacggtgacggtttt 1686

Qy   1286 aacccgctgaagaagcagggcgatgaagtcaaagcaggggagctgctgtgtgaattcgat 1345
Db   1687 gaagcaaaagttgctcaaggtaataaggtgaaagctggcgatgttcttggaacatttgac 1746
```

```
Qy   1346 attgatgccattaaggctgcaggttatgaggtaaccacgccgattgttgtttcgaat 1402
Db   1747 tcaaacaaaatcgctgcagctggacttgatgatacaacaatggttatcgttacaaat 1803
```

RESULT 11
AAQ55752/c
ID AAQ55752 standard; DNA; 3615 BP.
XX
AC AAQ55752;
XX
DT 25-SEP-1995 (first entry)
XX
DE Escherichia coli genomic probe EC-24.
XX
KW Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW K.pneumoniae; E.cloacae; clinical sample; ds.
XX
OS Escherichia coli.
XX
PN WO9401583-A.
XX
PD 20-JAN-1994.
XX
PF 07-JUL-1993; 93WO-JP00936.
XX
PR 07-JUL-1992; 92JP-0179719.
XX
PA (FUSO) FUSO PHARM IND LTD.
PA (OHNO/) OHNO T.
XX
PI Eda S, Matsuhisa A, Ohno T, Uehara H;
XX
DR WPI; 1994-035086/04.
XX
PT Probe for identifying bacteria causing infectious disease -
PT consists of a DNA fragment obtained by HindIII cleavage of the
PT pathogenic bacterial genomic DNA
XX
PS Claim 7; Page 62-64; 100pp; Japanese.
XX
CC The nucleotide sequence of a 3615 bp probe obtained by digestion of
CC Escherichia coli genomic DNA with the restriction enzyme HindIII.
CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,
CC by HindIII digestion of the genomes of Staphylococcus aureus,
CC S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used
CC to detect their respective microorganisms in clinical samples.
XX
SQ Sequence 3615 BP; 966 A; 962 C; 805 G; 882 T; 0 other;

Query Match 7.1%; Score 108; DB 15; Length 3615;
Best Local Similarity 46.2%; Pred. No. 2.8e-21;
Matches 539; Conservative 0; Mismatches 590; Indels 37; Gaps 4;

```
Qy    312 tcactccagtgctgacgttgctgctcaccggattccttacattcatcgccattggcccag 371
Db   2495 TCACACCATTGCTATGTCTGATGGTTATCACACCCGTCACCTTTCTGCTGGTGGGGCCGC 2436

Qy    372 caatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatgatttcggtg 431
Db   2435 TATCAACCTGGATAAGCGAACTGATTGCCGCCGGTTATCTCTGGCTTTATCAGGCGGTTC 2376

Qy    432 gtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcactggtctgc 491
Db   2375 CTGCATTTGCGGGCGCGGTAATGGGCGGCTTCTGGCAAATCTTCGTCATGTTCGGACTGC 2316

Qy    492 accagtccttcccgccaattgagctggagctgtttaaccagggtggatccttcatcttcg 551
Db   2315 ACTGGGGCCTGGTGCCGGTGTGTATCAATAACTTCACCGTGCTGGGCTACGACACCATGA 2256

Qy    552 caacggcatctatg---gctaatatcgcccagggtgcggcatgtttggcagtgttcttcc 608
Db   2255 TCCCGCTGTTAATGCCCGCCATTATGGCGCAGGTCGGGGCGGCGCTCGGCGTCTTCCTCT 2196

Qy    609 tggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttg 668
Db   2195 GCGAACGCGATGCGCAGAAAAAAGTGGTGGCGGGATCAGCGGCGTTGACGAGTCTGTTTG 2136

Qy    669 gtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcg 728
Db   2135 GTATCACCGAACCAGCGGTATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076

Qy    729 gtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaaggcagttg 788
Db   2075 CCTGTATCAGTGGGGCTTTGGGGGCCACCATTATTGGCTACGCGCAAACGAAAGTCTACT 2016

Qy    789 cgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatatggtcatgt 848
Db   2015 CCTTTGGTTTGCCAAGTATTTTCACCTTCATGCAAACCATCCCGTGAACGGGAATTGATT 1956

Qy    849 tcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgcttatggcc 908
Db   1955 TCACCGTCTGGGCCAGCGTTATTGGCGGTGTCATTGCCATCGGTTGCGCATTTGTCGGTA 1896

Qy    909 tttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctccagtgcctg 968
Db   1895 CGGTGATGCTTCATTTCATCACCGCTAAACGTCAGCCAGCGCAGGGTGCCCC-------- 1842

Qy    969 caggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattccaccatca 1028
Db   1843 --------------GAAGAGAAAACACCAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798

Qy   1029 tccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttg 1088
Db   1797 TCTGTTCACCGATGACGGGAGAGATTGTGTCGCTCATTCACGTCGCTGATACCACGTTTG 1738

Qy   1089 ccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctc 1148
Db   1737 CCAGTGGCCTGTTGGGTAAAGGTATTGCCATTCTGCCCTCGGTTGGTGAAGTGCGTTCTC 1678

Qy   1149 cggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagttcgcacca 1208
Db   1677 CGGTTGCGGGTCGAATTGCTTCGTTGTTCGCCACATTACA------CGCCATTGGCATTG 1624

Qy   1209 aggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacc 1268
Db   1623 AGTCAGATGATGG------TGTGGAGATCCTGATTCATGTCGGTATCGACACCGTAAAAC 1570

Qy   1269 tcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagc 1328
Db   1569 TGGACGGCAAATTCTTTTCCGCTCACGTCAACGTGGGTGACAAGGTCAATACAGGCGATC 1510

Qy   1329 tgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaaccacgccga 1388
Db   1509 GGCTGATTTCTTTTGATATCCCTGCTATTCGCGAGGCCGGATTTGATCTGACGACGCCGG 1450

Qy   1389 ttgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattg 1448
Db   1449 TATTAATCAGTAATAGCGATGATTTTACGGACGTATTACCCCACGGCACGGCGCAGATAA 1390

Qy   1449 aagcgggagccaacctgctcaacgtc 1474
Db   1389 GCGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 12
AAV52163/c
ID AAV52163 standard; DNA; 9769 BP.
XX
AC AAV52163;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:30.
```

```
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 318-324; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 9769 BP; 2918 A; 2141 C; 1792 G; 2916 T; 2 other;


  Query Match             7.1%;  Score 107.8;  DB 19;  Length 9769;
  Best Local Similarity   46.6%;  Pred. No. 5.5e-21;
  Matches  552;  Conservative    0;  Mismatches  597;  Indels   36;  Gaps    5;

Qy    302 gacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcc 361
Db   8979 GACCTCTTGGTAACACCATTCGTGACACTTTTGGTCATGTCTATCCTTGGACTCTTTGTC 8920

Qy    362 attggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttat 421
Db   8919 ATTGGACCAGTTTTCCACGTTGTTGAAAACTACATCCTTATTGCTACAAAAGCGATTCTT 8860

Qy    422 gatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatc 481
Db   8859 AGCATGCCATTTGGTCTTGGTGGTTTCTTGATTGGTGGGGTTCACCAATTGATCGTCGTG 8800

Qy    482 actggtctgcaccagtccttcccgccaattgagctggagctgtttaaccaggg---tgga 538
Db   8799 TCAGGTGTGCACCACATCTTCAACTTGCTTGAAGTGCAATTACTTGCTGCTGACCATGCT 8740

Qy    539 tccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggca 598
Db   8739 AACCCATTCAACGCTATCATCACAGCTGCTATGACAGCTCAAGGTGCTGCTACTGTTGCG 8680

Qy    599 gtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctcc 658
Db   8679 GTTGGTGTTAAAACAAAAAATCCAAAACTGAAAACACTTGCTTTCCCGGCTGCTCTTTCT 8620

Qy    659 gctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccg 718
Db   8619 GCCTTCCTAGGTATTACAGAGCCTGCTATCTTCGGGGTGAACTTGCGCTTCCGTAAACCA 8560

Qy    719 ttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatc 778
Db   8559 TTCTTCCTTTCATTGATTGCTGGTGCAATCGGTGGTGGATTGGCTTCTATCCTTGGACTT 8500

Qy    779 aaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagat 838
Db   8499 GCTGGTACTGGTAATGGTATCACCATCATCCCTGGTACAATGCTTTATGTTGGTAACGGA 8440

Qy    839 atggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgatt 898
Db   8439 CAACTTCCACAATACCTTCTTATGGTAGCTGTATCATT--------------TGCCCTT 8395

Qy    899 gcttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgct 958
Db   8394 GGTTTTGCTCTTACTTACATGTTTGGTTACG---AAGATGAAGTAGACGCAACTGCAGCT 8338

Qy    959 ccagtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgat 1018
Db   8337 GCAAAACGAGCTGAAGTGGCTGAAGAAAAAGAAGAAGTTGCGCCAGCAGCTCTTCAAAAT 8278

Qy   1019 tccaccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgat 1078
Db   8277 GAAACACTTGT---AACTCCTATCGTCGGTGATGTTGTCGCTCTTGCTGATGTCAATGAC 8221

Qy   1079 gccatgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcag 1138
Db   8220 CCAGTCTTCTCAAGTGGAGCTATGGGACAAGGTATCGTTGTGAAACCAAGCCAAGGCGTG 8161

Qy   1139 ttagtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgca 1198
Db   8160 GTCTATGCACCAGCTGATGCTGAAGTTTCAATTGCCTTTCCAACAGGGCACGCTTTTGGT 8101

Qy   1199 gttcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgac 1258
Db   8100 TTGAAAACAAGAAATGGTGCT-----------GAAGTTTTGATTCATGTTGGTATTGAT 8053

Qy   1259 acagtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaa 1318
Db   8052 ACTGTATCTATGAACGGTGACGGTTTTGAAACAAAAGTTGCTCAAGGTAATAAGGTGAAA 7993

Qy   1319 gcaggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggta 1378
Db   7992 GCTGGCGATGTTCTTGGAACATTTGACTCAAACAAAATCGCTGCAGCTGGACTTGATGAT 7933

Qy   1379 accacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttg 1438
Db   7932 ACAACAATGGTTATCGTTACAAATACAGGTGACTACGCTTCAGTAGCTCCAGTCGCAACA 7873

Qy   1439 ggcgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaa 1483
Db   7872 GGTTCAGTTGCTAAGGGGGATGCTGTGATCGAAGTGAAAATCTAA 7828


RESULT  13
AAT47502/c
ID   AAT47502 standard; DNA; 465 BP.
XX
AC   AAT47502;
XX
DT   19-FEB-1998  (first entry)
```

```
XX
DE   Partial P. rubrum scrA gene region.
XX
KW   scr operon; scrA; sucrose metabolising enzyme; sucrose isomerase;
KW   reduced metabolism; non-cariogenic sugar; trehalulose; palatinose;
KW   decomposition; monosaccharide; improved yield; disaccharide; ds.
XX
OS   Protaminobacter rubrum.
XX
PN   EP751218-A2.
XX
PD   02-JAN-1997.
XX
PF   27-JUN-1996;   96EP-0110396.
XX
PR   28-JUN-1995;   95DE-4023560.
XX
PA   (SUED-) SUEDZUCKER AG MANNHEIM/OCHSENFURT.
XX
PI   Klein K,  Mattes R,  Stegmaier S;
XX
DR   WPI; 1997-054676/06.
XX
PT   Cell with reduced sucrose metabolism but with sucrose isomerase
PT   activity - providing high yields of non-cariogenic sugars, esp.
PT   trehalulose and palatinose
XX
PS   Claim 4; Pages 12-13; 26pp; German.
XX
CC   The present sequence comprises a partial DNA sequence from the scr operon
CC   containing the scrA gene region. The scr operon encodes sucrose
CC   metabolising enzymes, in particular sucrose isomerase. Cells containing
CC   at least one DNA sequence encoding a protein with sucrose isomerase
CC   activity, and having reduced ability to metabolise sucrose, or their
CC   extracts, optionally in an immobilised form, are used to produce
CC   non-cariogenic sugars, especially trehalulose and/or palatinose. During
CC   production of the non-cariogenic sugars, decomposition of sucrose to
CC   monosaccharides (which are difficult to separate from the product) is
CC   minimised, so the yield of the disaccharides is improved.
XX
SQ   Sequence 465 BP; 106 A; 137 C; 149 G; 73 T; 0 other;


  Query Match             6.4%;  Score 97.8;  DB 18;  Length 465;
  Best Local Similarity   53.5%;  Pred. No. 1e-18;
  Matches  204;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy    145 cgccaccatggctgcgggcgaaatgccaatgtggtccctgtttggtttagatgttgccca 204
          ||||      |  |||||||   | |  |     |    | || ||  | ||| | |||
Db    441 CGCCTGGGGCGTCGCGGGCGGTTTCCACACCATGAATTTCTTCGGACTCGATATCGCCAT 382

Qy    205 agccggttaccagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgat 264
            ||| || ||||| || ||| | || ||||||| ||      ||| || ||     | |
Db    381 GATCGGCTATCAGGGTACGGTGTTCCCGGTGCTGCTGACGGTGTGGTTTATGAGCGTGCT 322

Qy    265 cgagaagttcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgct 324
           || ||    |||| ||      ||     | |    ||| |  |  | || ||| | ||
Db    321 GGAAAAACAGCTGCGCAGAGTGATCCCCAACGCGCTGGACCTGATCCTGACGCCATTCCT 262

Qy    325 gacgttgctgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggt 384
          |||| || || || |||| ||| |  |  | |    || ||||| ||   ||||    |
Db    261 GACGGTGGTGATCTCCGGCTTCGTCGCCATGCTGTTTATCGGCCCCGCCGGGCGCACACT 202

Qy    385 gggcgatgtgctggcacacggtctacagggactttatgatttcggtggtccagtcggcgg 444
          ||| || |   |  |   ||  ||       ||   ||    ||  |||    ||| |||
Db    201 GGGTGACGGCATCTCCTTCGTGCTCAGCACGCTGATTGCCCACGCCGGTTGGTTCGCCGG 142

Qy    445 tctgctcttcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttccc 504
            | || |||||     | || ||  | ||||||||||| ||  | |||||   |||||
Db    141 ATTCCTGTTCGGCGGCCTGTATTCCGCCATCGTCATCACCGGCATTCACCACAGCTTCCA 82

Qy    505 gccaattgagctggagctgtt 525
            |  | |||  || |||| |
Db     81 CGCCGTAGAGGGGGGGCTGCT 61


RESULT  14
AAV74367
ID   AAV74367 standard; DNA; 30246 BP.
XX
AC   AAV74367;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #56.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1741..1800
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    3541..3600
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    5341..5400
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    7141..7200
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    8941..9000
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    10741..10800
FT                   /*tag= f
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    12541..12600
FT                   /*tag= g
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    14341..14400
FT                   /*tag= h
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    16141..16200
FT                   /*tag= i
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
```

```
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    17941..18000
FT                   /*tag= j
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    19741..19800
FT                   /*tag= k
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    21541..21600
FT                   /*tag= l
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    23341..23400
FT                   /*tag= m
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    25141..25200
FT                   /*tag= n
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    26941..27000
FT                   /*tag= o
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
FT   misc_feature    28741..28800
FT                   /*tag= p
FT                   /note= "these bases represent a line of missing text in
FT                      the sequence listing in the specification. They
FT                      are included to maintain the nucleotide numbering
FT                      given in the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-0100117.
XX
PR   05-JAN-1996;   96US-0009861.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Fannon MR,  Kunsch CA;
PI   Rosen CA;
XX
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
XX
PS   Claim 1; Page 422-438; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
XX
SQ   Sequence 30246 BP; 9663 A; 4436 C; 6386 G; 8791 T; 970 other;


  Query Match             5.6%;  Score 86.2;  DB 18;  Length 30246;
  Best Local Similarity   48.5%;  Pred. No. 2.7e-14;
  Matches  309;  Conservative    0;  Mismatches  313;  Indels   15;  Gaps    2;

Qy     153 tggctgcgggcgaaatgccaatgtggtccctgtttggtttagatgttgcccaagccggtt 212
           ||||    ||  | || || | ||||  | | |||||  |||||  ||   ||      ||
Db   11413 tggcaaaagggaatattccgacgtggaacttatttggcttagagattaagcagttgaatt 11472

Qy     213 accagggcaccgtgcttcctgtgctggtggtttcttggattctggcaacgatcgagaagt 272
           ||||  ||    |||  | ||  ||  |  | |   |||   |||| || |  || ||  ||
Db   11473 accaaggtcaagtgttgccagttttaattgcagcttacgttctagctaaaattgaaaaag 11532

Qy     273 tcctgcacaagcgactcaagggcactgcagacttcctgatcactccagtgctgacgttgc 332
             |  | ||     |  | |   |   | |  |  || ||  |  |    |  || |
Db   11533 gattaaataaagtcgttcacgattcgataaaaatgttggtcgttggacccgtagcgcttt 11592

Qy     333 tgctcaccggattccttacattcatcgccattggcccagcaatgcgctgggtgggcgatg 392
           |  | || |||||  |  |||| ||   |||||| ||||    |   | | | ||    |
Db   11593 tagttactggatttttagcatttattatcattggaccagttgcgttattgattggtacag 11652

Qy     393 tgctggcacacggtctacagggacttttatgatttcggtggtccagtcggcggtctgctct 452
             |  ||   ||| |         | |   |    |  ||     | |||||     | |
Db   11653 gtattacatctggtgttacatttatattccaacatgcaggatggcttggcggagcaatat 11712

Qy     453 tcggtctggtctactcaccaatcgtcatcactggtctgcaccagtccttcccgccaattg 512
             ||  || | ||  ||||| | || || || || || ||||||    ||     || | |
Db   11713 atggattgttatatgcaccacttgtaattacaggactacaccatatgtttttagcagtag 11772

Qy     513 agctggagctgtt------taaccagggtggatccttcatcttcgcaacggcatctatgg 566
           |  |  |  || |       | |   || ||  | |   | |   |||  |   | ||
Db   11773 atttccaattgatgggtagcagcttaggcggtacgtatttatggccaattgttgcgattt 11832

Qy     567 ctaatatcgcccagggtgcggcatgtttggcagtgttcttcctggcgaagagt------- 619
           | |||||    |||||  | |||   || | ||  |  ||  |    ||  ||
Db   11833 ccaatatttgtcagggctctgcagcatttggagcatggtttgtctataaacgtcgtaaaa 11892

Qy     620 --gaaaagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttggtattacgg 677
             |  ||     | ||| | |||    | ||  || | || | | |  | ||| |||| |
Db   11893 tggttaaagaagaaggcttggcattaacatcttgtatttctggtatgttaggtgttactg 11952

Qy     678 agcctgcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggta 737
           | || || || ||||||||||||| | |  |||   |  || ||  ||   |  ||    |
Db   11953 aaccagccatgttcggtgtgaacttacctctgaaatatccatttatcgctgcgatatcaa 12012

Qy     738 ccgcagctatcggtggcgctttgattgcactctttaa 774
           |  |   | |    || ||  |  |||   |   |||
Db   12013 cgtcttgtgtattgggggcaatcgttggtatgaataa 12049


RESULT  15
ABA03041/c
ID   ABA03041 standard; DNA; 2944528 BP.
XX
AC   ABA03041;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes EGD-e genome sequence.
```

```
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease; ds.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C,  Frangeul L,  Couve E,  Rusniok C,  Fsihi H,  Dehoux P;
PI   Dussurget O,  Chetouani F,  Nedjari H,  Glaser P,  Kunst F,  Cossart P;
PI   Daniels J,  Goebel W,  Kreft J,  Kuhn M,  Ng E,  Vazquez-Boland JA;
PI   Dominguez-Bernal G,  Garrido-Garcia P,  Tierrez-Martinez A,  Amend A;
PI   Chakraborty T,  Domann E,  Hain T,  Berche P,  Charbit A,  Durant L;
PI   Perez-Diaz J,  Baquero F,  Garcia Del Portillo F,  Gomez-Lopez N;
PI   Maduenio E,  De Pablos B,  Wehland J,  Kaerst U,  Entian K,  Hauf J;
PI   Rose M,  Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides  -
XX
PS   Claim 1; SEQ ID No 1; 192pp; French.
XX
CC   The present sequence is the genome sequence of Listeria monocytogenes
CC   EGD-e. This sequence and fragments of this sequence are useful for
CC   selecting probes and primers for detecting genes in L. monocytogenes and
CC   related organisms, and to study genetic polymorphisms and other genomes.
CC   Proteins (ABB47297-ABB50149) expressed from the present sequence are
CC   useful for raising specific antibodies, identification of L.
CC   monocytogenes and related organisms, and for biosynthesis and
CC   biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC   proteins encoded by it are also useful for selecting compounds that
CC   regulate gene expression and cell replication and modulate L.
CC   monocytogenes-related diseases. In addition, this sequence and proteins
CC   encoded by it are useful in pharmaceutical and vaccines compositions for
CC   the treatment or prevention of infections by L. monocytogenes and related
CC   organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;


  Query Match             5.3%;  Score 81.6;  DB 24;  Length 2944528;
  Best Local Similarity   46.1%;  Pred. No. 7.6e-12;
  Matches  310;  Conservative    0;  Mismatches  359;  Indels    3;  Gaps    1;

Qy    171 caatgtggtccctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttc 230
          ||| |     | |  | || |||   ||  | | ||     ||       | | ||  |||
Db 2855687 CAACGCCACACTTTATCGGGTTACCAGTAACTCCAGTGACGTATGCTTACTCTGTTATTC 2855628

Qy    231 ctgtgctggtggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactca 290
          |  |  |  | |    ||||||| || |      | || || |   || |   |    |
Db 2855627 CGATTTTACTTGCAATTTGGATTATGTCTTATGTGGAAAAATGGGTGGATCGGATTATTC 2855568

Qy    291 agggcactgcagacttcctgatcactccagtgctgacgttgctgctcaccggattcctta 350
            | |   | | || ||| |    ||| |  | ||  || |  ||    |      | |
Db 2855567 CAACCTCGTTAAAATTGCTGTTTGTCCCATTAATTACAATGTTTATCGTTGTTCCTTTGA 2855508

Qy    351 cattcatcgccattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctac 410
          |  |  | |- ||||| |||  |       |  || || || |   |  |   |||  |
Db 2855507 CGCTTGTTGTTATTGGACCACTAGGTACATTTGTTGGAGACGGTGTATCTGGCGGAATTA 2855448

Qy    411 agggactttatgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcac 470
          |  |  | |   ||   |||||  |  | || ||| |  |  | |||   | |    |
Db 2855447 ACTGGTTATTAAATAATGGTGGAGCTATTGGTGGTATCTTGATTGGTGGTGCCATGGCCA 2855388

Qy    471 caatcgtcatcactggtctgcaccagtccttcccgccaattgagctggagctgtttaacc 530
           ||||| || |||||| ||||| |  |  |    ||| |||   |       ||
Db 2855387 TTATCGTTATGACTGGTATGCACTATGCAATTGTTCCATTTGTTATTAGTAATTTAGCAA 2855328

Qy    531 agggtggatccttcatcttcgcaacggcatc---tatggctaatatcgcccagggtgcgg 587
          |  ||| |     |  ||   | ||  | |   |||  | |||||   ||| |  |  |
Db 2855327 AATATGGTTATGATAAATTTTTACCGTTAACTTATATTTCCAATATGAGCCAAGCGGGAG 2855268

Qy    588 catgtttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgctt 647
          |   ||| |  || || ||||  || ||   | |||| || ||     | |||  | |
Db 2855267 CGACTTTCGGTGTATTTTTCCGAGCAAAAGATAAAAAACTGAAATCATTAGCATTTTCAA 2855208

Qy    648 caggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcc 707
          | ||| |  | ||  |  | ||  | ||||| || ||||| | |||  | ||| |
Db 2855207 CTGGTTTAACTGCATTAATGGGCGTAACGGAACCAGCGATGTACGGGATAAACGTGGTAT 2855148

Qy    708 tgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcac 767
              | || ||| |      |   |   |  |||   |||||||  ||
Db 2855147 ATAAACGTCCTTTCATGGCATCGTTAATTGGTGGGGCTGCTGGTGGCGGCTTTGCAATGA 2855088

Qy    768 tctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattg 827
          | |||  | | || ||   ||  ||    | |   |||    |    |   || |   |
Db 2855087 TGTTTGGTGTAAAAGCGTATGTCTTAACTGGTAACGGTGGTATTCCAGGACTTCCAGGTC 2855028

Qy    828 atgctccagata 839
            ||    |||||
Db 2855027 TTGTCGGAGATA 2855016


Search completed: September 25, 2002, 09:14:10
Job time: 10631 sec
```

GenCore version 4.5 .

OM nucleic - nucleic search, using sw model

Run on: September 25, 2002, 06:05:49 ; Search time 2033.44 Seconds
(without alignments)
15714.662 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 ctcatggcatctgcgccgtt.....gttgaaaccttgagtgttcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1527	100.0	1527	6	AX069134	AX069134 Sequence
c 2	1515	99.2	349980	6	AX127152	AX127152 Sequence
3	1492	97.7	1983	6	AX122988	AX122988 Sequence
4	1107.4	72.5	1109	6	AX069136	AX069136 Sequence
c 5	287.6	18.8	21838	1	PDCRAFOPER	L32093 Pediococcus
c 6	287.6	18.8	21839	1	PPSURFOP	Z32771 P.pentosace
7	253.4	16.6	2788	1	AF401046	AF401046 Lactobaci
8	250.8	16.4	5800	1	LLZ97015	Z97015 Lactococcus
c 9	231.8	15.2	10264	1	AE004395	AE004395 Vibrio ch
10	230.8	15.1	2508	1	STRSCRA	M22711 Streptococc
c 11	218.6	14.3	10085	1	AE006222	AE006222 Pasteurel
12	215.2	14.1	2955	1	VIBSCRAK	M76768 Vibrio algi
13	210.6	13.8	2655	1	SXSCRA	X69800 S.xylosus s
14	205.4	13.5	4158	1	BSTREAPR	Z54245 B.subtilis
c 15	201.8	13.2	22197	1	D83967	D83967 Bacillus su
c 16	201.8	13.2	37900	1	D86417	D86417 Bacillus su
17	201.8	13.2	208430	1	BSUB0005	Z99108 Bacillus su
c 18	194.8	12.8	341350	1	AP003365	AP003365 Staphyloc
c 19	190.6	12.5	2913	1	AF269422	AF269422 Staphyloc
c 20	190.6	12.5	2913	6	AX144742	AX144742 Sequence
c 21	183	12.0	299850	1	AP001514	AP001514 Bacillus
c 22	174	11.4	215050	1	AL646084	AL646084 Ralstonia
23	157.6	10.3	4487	1	AF229829	AF229829 Pseudomon
24	146.8	9.6	292550	1	AP001513	AP001513 Bacillus
25	146.4	9.6	296950	1	AP001508	AP001508 Bacillus
26	143.6	9.4	4621	1	KPSCRYAB	X57401 K.pneumonia
27	141.8	9.3	6950	1	EAM250722	AJ250722 Erwinia a
28	138	9.0	5065	1	ERWBGPA	M81772 Erwinia chr
29	136.2	8.9	1398	1	BSU34876	U34876 Bacillus st
c 30	135.2	8.9	2819	1	BACISPQ	D37921 Alkalophiol
31	134.6	8.8	4322	1	STSCRCOMP	X67750 S.typhimuri
32	134.6	8.8	231450	1	AL596163	AL596163 Listeria
33	132.2	8.7	3518	1	BACSACP	J03006 B.subtilis
34	132.2	8.7	97015	1	BSGENR	X73124 B.subtilis
35	132.2	8.7	200050	1	AL591973	AL591973 Listeria
c 36	132.2	8.7	212150	1	BSUB0020	Z99123 Bacillus su
37	130	8.5	1530	1	STSCRA	Y00541 Salmonella
c 38	126.6	8.3	3895	6	BD003881	BD003881 Polynucle
c 39	126.6	8.3	13589	1	AE008535	AE008535 Streptoco
c 40	126.6	8.3	14515	1	AE007479	AE007479 Streptoco
c 41	123.4	8.1	151947	2	SPNEU1902	AL449924 Streptoco
42	120.8	7.9	8006	1	AF206272	AF206272 Streptoco
43	117.8	7.7	5600	1	AF205034	AF205034 Clostridi
44	117.8	7.7	10611	1	AE007557	AE007557 Clostridi
c 45	114.2	7.5	136254	1	ECOUW82	L10328 E. coli; th

ALIGNMENTS

RESULT 1
AX069134
LOCUS AX069134 1527 bp DNA linear PAT 25-JAN-2001
DEFINITION Sequence 1 from Patent WO0102583.
ACCESSION AX069134
VERSION AX069134.1 GI:12579016
KEYWORDS .
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 1527)
AUTHORS Pompejus,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
TITLE Orynebacterium glutamicum genes encoding phosphoenolpyruvat e:
sugar phosphotransferase system proteins
JOURNAL Patent: WO 0102583-A 1 11-JAN-2001;
BASF AKTIENGESELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1. .1527
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
CDS 101. .1507

```
                     /note="RXS00315"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC27276.1"
                     /db_xref="GI:12579017"
                     /translation="MAMVFPSLVNGYDVAATMAAGEMPMWSLFGLDVAQAGYQGTVLP
                     VLVVSWILATIEKFLHKRLKGTADFLITPVLTLLLTGFLTFIAIGPAMRWVGDVLAHG
                     LQGLYDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQGGSFIFATASMANIAQ
                     GAACLAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFFIGIGTAAIGG
                     ALIALFNIKAVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRRN
                     GSIDPDATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKL
                     GSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMHIGFDTVNLNG
                     THFNPLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIE
                     AGANLLNVAKKEAVPATP"
BASE COUNT      304 a    392 c    430 g    401 t
ORIGIN


  Query Match           100.0%;  Score 1527;  DB 6;  Length 1527;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1527;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCATGGCATCTGCGCCGTTCGCGTTCTTGCCAGTGTTGGTTGGTTTCACCGCAACCAAG 60

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGTTTCGGCGGCAATGAGTTCCTGGGCGCCGCGTATTGGTATGGCGATGGTGTTCCCGAG 120

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTGGTGAACGGCTACGACGTGGCCGCCACCATGGCTGCGGGCGAAATGCCAATGTGGTC 180

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCTGTTTGGTTTAGATGTTGCCCAAGCCGGTTACCAGGGCACCGTGCTTCCTGTGCTGGT 240

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGTTTCTTGGATTCTGGCAACGATCGAGAAGTTCCTGCACAAGCGACTCAAGGGCACTGC 300

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGACTTCCTGATCACTCCAGTGCTGACGTTGCTGCTCACCGGATTCCTTACATTCATCGC 360

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATTGGCCCAGCAATGCGCTGGGTGGGCGATGTGCTGGCACACGGTCTACAGGGACTTTA 420

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGATTTCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGGTCTACTCACCAATCGTCAT 480

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CACTGGTCTGCACCAGTCCTTCCCGCCAATTGAGCTGGAGCTGTTTAACCAGGGTGGATC 540

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCCAGGGTGCGGCATGTTTGGCAGT 600

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTTCTTCCTGGCGAAGAGTGAAAAGCTCAAGGGCCTTGCAGGTGCTTCAGGTGTCTCCGC 660

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TGTTCTTGGTATTACGGAGCCTGCGATCTTCGGTGTGAACCTTCGCCTGCGCTGGCCGTT 720

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATCAA 780

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGCAGTTGCGTTGGGCGCTGCAGGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGATAT 840

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATTGC 900

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTATGGCCTTTACTTGGTTCGCCGCAACGGCAGCATTGATCCAGATGCAACCGCTGCTCC 960

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGTGCCTGCAGGAACGACCAAAGCCGAAGCAGAAGCACCCGCAGAATTTTCAAACGATTC 1020

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGATGC 1080

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CATGTTTGCCAGCGGAAAGCTTGGCTCGGGCGTTGCCATCGTCCCAACCAAGGGGCAGTT 1140

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AGTTTCTCCGGTGAGTGGAAAGATTGTGGTGGCATTCCCATCTGGCCATGCTTTCGCAGT 1200

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGCACATTGGTTTCGACAC 1260

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGTAAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAAGC 1320

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAGGCTGCAGGTTATGAGGTAAC 1380

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CACGCCGATTGTTGTTTCGAATTACAAGAAAACCGGACCTGTAAACACTTACGGTTTGGG 1440

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CGAAATTGAAGCGGGAGCCAACCTGCTCAACGTCGCAAAGAAAGAAGCGGTGCCAGCAAC 1500

Qy   1501 accataagttgaaaccttgagtgttcg 1527
          |||||||||||||||||||||||||||
Db   1501 ACCATAAGTTGAAACCTTGAGTGTTCG 1527


RESULT   2
AX127152/c
LOCUS       AX127152               349980 bp    DNA     linear   PAT 11-MAY-2001
DEFINITION  Sequence 7068 from Patent EP1108790.
ACCESSION   AX127152 AX114121
VERSION     AX127152.1  GI:14041140
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 7068 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
```

```
     source          1. .349980
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
                     /note="Seq 1 to long (3.309.400) split in 11, seq 7068
                     2.700.001 3.049.980"
BASE COUNT    81250 a  97718 c  90621 g  80391 t
ORIGIN


  Query Match             99.2%;  Score 1515;  DB 6;  Length 349980;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1526;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111326 CTCATGGCATCTGCGCCGTTCGCGTTCTTGCCAGTGTTGGTTGGTTTCACCGCAACCAAG 111267

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 111266 CGTTTCGGCGGCAATGAGTTCCTGGGCGCCG-GTATTGGTATGGCGATGGTGTTCCCGAG 111208

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111207 CTTGGTGAACGGCTACGACGTGGCCGCCACCATGGCTGCGGGCGAAATGCCAATGTGGTC 111148

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111147 CCTGTTTGGTTTAGATGTTGCCCAAGCCGGTTACCAGGGCACCGTGCTTCCTGTGCTGGT 111088

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111087 GGTTTCTTGGATTCTGGCAACGATCGAGAAGTTCCTGCACAAGCGACTCAAGGGCACTGC 111028

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111027 AGACTTCCTGATCACTCCAGTGCTGACGTTGCTGCTCACCGGATTCCTTACATTCATCGC 110968

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110967 CATTGGCCCAGCAATGCGCTGGGTGGGCGATGTGCTGGCACACGGTCTACAGGGACTTTA 110908

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110907 TGATTTCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGGTCTACTCACCAATCGTCAT 110848

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110847 CACTGGTCTGCACCAGTCCTTCCCGCCAATTGAGCTGGAGCTGTTTAACCAGGGTGGATC 110788

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110787 CTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCCAGGGTGCGGCATGTTTGGCAGT 110728

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110727 GTTCTTCCTGGCGAAGAGTGAAAAGCTCAAGGGCCTTGCAGGTGCTTCAGGTGTCTCCGC 110668

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110667 TGTTCTTGGTATTACGGAGCCTGCGATCTTCGGTGTGAACCTTCGCCTGCGCTGGCCGTT 110608

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110607 CTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATCAA 110548

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110547 GGCAGTTGCGTTGGGCGCTGCAGGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGATAT 110488

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110487 GGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATTGC 110428

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110427 TTATGGCCTTTACTTGGTTCGCCGCAACGGCAGCATTGATCCAGATGCAACCGCTGCTCC 110368

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaatttttcaaacgattc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110367 AGTGCCTGCAGGAACGACCAAAGCCGAAGCAGAAGCACCCGCAGAATTTTCAAACGATTC 110308

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110307 CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGATGC 110248

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110247 CATGTTTGCCAGCGGAAAGCTTGGCTCGGGCGTTGCCATCGTCCCAACCAAGGGGCAGTT 110188

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110187 AGTTTCTCCGGTGAGTGGAAAGATTGTGGTGGCATTCCCATCTGGCCATGCTTTCGCAGT 110128

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110127 TCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGCACATTGGTTTCGACAC 110068

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110067 AGTAAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAAGC 110008

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110007 AGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAGGCTGCAGGTTATGAGGTAAC 109948

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109947 CACGCCGATTGTTGTTTCGAATTACAAGAAAACCGGACCTGTAAACACTTACGGTTTGGG 109888

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109887 CGAAATTGAAGCGGGAGCCAACCTGCTCAACGTCGCAAAGAAAGAAGCGGTGCCAGCAAC 109828

Qy   1501 accataagttgaaaccttgagtgttcg 1527
          |||||||||||||||||||||||||||
Db 109827 ACCATAAGTTGAAACCTTGAGTGTTCG 109801


RESULT   3
AX122988
LOCUS       AX122988               1983 bp    DNA     linear   PAT 11-MAY-2001
DEFINITION  Sequence 2904 from Patent EP1108790.
ACCESSION   AX122988
VERSION     AX122988.1  GI:14040476
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 1983)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 2904 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
     source          1. .1983
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
BASE COUNT      410 a    508 c    558 g    507 t
ORIGIN


  Query Match             97.7%;  Score 1492;  DB 6;  Length 1983;
  Best Local Similarity   99.9%;  Pred. No. 0;
```

```
  Matches 1503;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTCATGGCATCTGCGCCGTTCGCGTTCTTGCCAGTGTTGGTTGGTTTCACCGCAACCAAG 540

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    541 CGTTTCGGCGGCAATGAGTTCCTGGGCGCCG-GTATTGGTATGGCGATGGTGTTCCCGAG 599

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 CTTGGTGAACGGCTACGACGTGGCCGCCACCATGGCTGCGGGCGAAATGCCAATGTGGTC 659

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 CCTGTTTGGTTTAGATGTTGCCCAAGCCGGTTACCAGGGCACCGTGCTTCCTGTGCTGGT 719

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 GGTTTCTTGGATTCTGGCAACGATCGAGAAGTTCCTGCACAAGCGACTCAAGGGCACTGC 779

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 AGACTTCCTGATCACTCCAGTGCTGACGTTGCTGCTCACCGGATTCCTTACATTCATCGC 839

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 CATTGGCCCAGCAATGCGCTGGGTGGGCGATGTGCTGGCACACGGTCTACAGGGACTTTA 899

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 TGATTTCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGGTCTACTCACCAATCGTCAT 959

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtggatc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 CACTGGTCTGCACCAGTCCTTCCCGCCAATTGAGCTGGAGCTGTTTAACCAGGGTGGATC 1019

Qy    541 cttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggcagt 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 CTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCCAGGGTGCGGCATGTTTGGCAGT 1079

Qy    601 gttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctccgc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 GTTCTTCCTGGCGAAGAGTGAAAAGCTCAAGGGCCTTGCAGGTGCTTCAGGTGTCTCCGC 1139

Qy    661 tgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccgtt 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 TGTTCTTGGTATTACGGAGCCTGCGATCTTCGGTGTGAACCTTCGCCTGCGCTGGCCGTT 1199

Qy    721 cttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 CTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATCAA 1259

Qy    781 ggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagatat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 GGCAGTTGCGTTGGGCGCTGCAGGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGATAT 1319

Qy    841 ggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgattgc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1320 GGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATTGC 1379

Qy    901 ttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgctcc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 TTATGGCCTTTACTTGGTTCGCCGCAACGGCAGCATTGATCCAGATGCAACCGCTGCTCC 1439

Qy    961 agtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgattc 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 AGTGCCTGCAGGAACGACCAAAGCCGAAGCAGAAGCACCCGCAGAATTTTCAAACGATTC 1499

Qy   1021 caccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgatgc 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500 CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGATGC 1559

Qy   1081 catgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcagtt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 CATGTTTGCCAGCGGAAAGCTTGGCTCGGGCGTTGCCATCGTCCCAACCAAGGGGCAGTT 1619

Qy   1141 agtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 AGTTTCTCCGGTGAGTGGAAAGATTGTGGTGGCATTCCCATCTGGCCATGCTTTCGCAGT 1679

Qy   1201 tcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgacac 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 TCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGCACATTGGTTTCGACAC 1739

Qy   1261 agtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaagc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 AGTAAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAAGC 1799

Qy   1321 aggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggtaac 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 AGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAGGCTGCAGGTTATGAGGTAAC 1859

Qy   1381 cacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttggg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 CACGCCGATTGTTGTTTCGAATTACAAGAAAACCGGACCTGTAAACACTTACGGTTTGGG 1919

Qy   1441 cgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagcaac 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 CGAAATTGAAGCGGGAGCCAACCTGCTCAACGTCGCAAAGAAAGAAGCGGTGCCAGCAAC 1979

Qy   1501 acca 1504
          ||||
Db   1980 ACCA 1983


RESULT   4
AX069136
LOCUS       AX069136               1109 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 3 from Patent WO0102583.
ACCESSION   AX069136
VERSION     AX069136.1  GI:12579018
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 1109)
  AUTHORS   Pompejus,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
  TITLE     Orynebacterium glutamicum genes encoding phosphoenolpyruvat e:
            sugar phosphotransferase system proteins
  JOURNAL   Patent: WO 0102583-A 3 11-JAN-2001;
            BASF AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1. .1109
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
     CDS             <1. .1089
                     /note="FRXA00315"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC27277.1"
                     /db_xref="GI:12579019"
                     /translation="YDFGGPVGGLLFGLVYSPIVITGLHQSFPPIELELFNQGGSFIF
                     ATASMANIAQGAACLAVFFLAKSEKLKGLAGASGVSAVLGITEPAIFGVNLRLRWPFF
                     IGIGTAAIGGALIALFNIKAVALGAAGFLGVVSIDAPDMVMFLVCAVVTFFIAFGAAI
                     AYGLYLVRRNGSIDPDATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSV
                     SDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFAVRTKAEDGSNVDILMH
                     IGFDTVNLNGTHFNPLKKQGDEVKAGELLCEFDIDAIKAAGYEVTTPIVVSNYKKTGP
                     VNTYGLGEIEAGANLLNVAKKEAVPATP"
BASE COUNT      236 a    278 c    303 g    292 t
ORIGIN
```

```
  Query Match           72.5%;  Score 1107.4;  DB 6;  Length 1109;
  Best Local Similarity  99.9%;  Pred. No. 1.3e-292;
  Matches 1108;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     419 tatgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtc 478
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TATGATTTCGGCGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGGTCTACTCACCAATCGTC 60

Qy     479 atcactggtctgcaccagtccttcccgccaattgagctggagctgtttaaccagggtgga 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATCACTGGTCTGCACCAGTCCTTCCCGCCAATTGAGCTGGAGCTGTTTAACCAGGGTGGA 120

Qy     539 tccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgtttggca 598
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCCTTCATCTTCGCAACGGCATCTATGGCTAATATCGCCCAGGGTGCGGCATGTTTGGCA 180

Qy     599 gtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgtctcc 658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTGTTCTTCCTGGCGAAGAGTGAAAAGCTCAAGGGCCTTGCAGGTGCTTCAGGTGTCTCC 240

Qy     659 gctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctggccg 718
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCTGTTCTTGGTATTACGGAGCCTGCGATCTTCGGTGTGAACCTTCGCCTGCGCTGGCCG 300

Qy     719 ttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaatatc 778
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTCTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATC 360

Qy     779 aaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctccagat 838
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAGGCAGTTGCGTTGGGCGCTGCAGGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGAT 420

Qy     839 atggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagcgatt 898
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATT 480

Qy     899 gcttatggcctttacttggttcgccgcaacggcagcattgatccagatgcaaccgctgct 958
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCTTATGGCCTTTACTTGGTTCGCCGCAACGGCAGCATTGATCCAGATGCAACCGCTGCT 540

Qy     959 ccagtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattttcaaacgat 1018
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CCAGTGCCTGCAGGAACGACCAAAGCCGAAGCAGAAGCACCCGCAGAATTTTCAAACGAT 600

Qy    1019 tccaccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcgat 1078
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCCACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGAT 660

Qy    1079 gccatgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaaccaaggggcag 1138
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCCATGTTTGCCAGCGGAAAGCTTGGCTCGGGCGTTGCCATCGTCCCAACCAAGGGGCAG 720

Qy    1139 ttagtttctccggtgagtggaaagattgtggtggcattcccatctggccatgctttcgca 1198
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TTAGTTTCTCCGGTGAGTGGAAAGATTGTGGTGGCATTCCCATCTGGCCATGCTTTCGCA 780

Qy    1199 gttcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacattggtttcgac 1258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTTCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGCACATTGGTTTCGAC 840

Qy    1259 acagtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgatgaagtcaaa 1318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ACAGTAAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAA 900

Qy    1319 gcaggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggttatgaggta 1378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GCAGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAGGCTGCAGGTTATGAGGTA 960

Qy    1379 accacgccgattgttgtttcgaattacaagaaaaccggacctgtaaacacttacggtttg 1438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 ACCACGCCGATTGTTGTTTCGAATTACAAGAAAACCGGACCTGTAAACACTTACGGTTTG 1020

Qy    1439 ggcgaaattgaagcgggagccaacctgctcaacgtcgcaaagaaagaagcggtgccagca 1498
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGCGAAATTGAAGCGGGAGCCAACCTGCTCAACGTCGCAAAGAAAGAAGCGGTGCCAGCA 1080

Qy    1499 acaccataagttgaaaccttgagtgttcg 1527
           |||||||||||||||||||||||||||||
Db    1081 ACACCATAAGTTGAAACCTTGAGTGTTCG 1109


RESULT   5
PDCRAFOPER/c
LOCUS       PDCRAFOPER              21838 bp    DNA     linear   BCT 23-MAY-1994
DEFINITION  Pediococcus pentosaceus raffinose operon genes.
ACCESSION   L32093
VERSION     L32093.1  GI:493181
KEYWORDS    agaR gene; agaS gene; agl gene; alpha-galactosidase;
            alpha-glucosidase; fructokinase; insertion element; permease; rafP
            gene; rafR gene; raffinose operon; regulatory protein; scrA gene;
            scrB gene; scrK gene; scrR gene; sucrase; sucrose-6-phosphate;
            transport protein.
SOURCE      Pediococcus pentosaceus (strain PPE1.0) DNA; Insertion sequence
            IS30 homolog (transposable element Insertion sequence IS30 homolog,
            kingdom Prokaryotae) DNA; Pediococcus pentosaceus (strain PPE1.0)
            DNA; Insertion sequence IS3 homolog (transposable element Insertion
            sequence IS3 homolog, kingdom Prokaryotae) DNA; and Pediococcus
            pentosaceus (strain PPE1.0) DNA.
  ORGANISM  Pediococcus pentosaceus
            Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
            Pediococcus.
REFERENCE   1  (bases 1 to 21838)
  AUTHORS   Leenhouts,K.J., Bolhuis,A.A., Kok,J.J. and Venema,G.G.
  TITLE     The sucrose and raffinose operons of Pediococcus pentosaceus PPE1.0
  JOURNAL   Unpublished (1994)
COMMENT     On May 25, 1994 this sequence version replaced gi:475106.
FEATURES             Location/Qualifiers
     source          1. .21838
                     /organism="Pediococcus pentosaceus"
                     /strain="PPE1.0"
                     /db_xref="taxon:1255"
     repeat_unit     487. .510
                     /note="DR2; putative"
                     /rpt_type=direct
     repeat_unit     488. .561
                     /note="IR1; putative"
                     /rpt_type=inverted
     repeat_unit     513. .561
                     /note="DR1; putative"
                     /rpt_type=direct
     terminator      2724. .2760
                     /standard_name="terminator 4"
                     /note="putative"
     gene            complement(2929. .3864)
                     /gene="rafR"
     CDS             complement(2929. .3762)
                     /gene="rafR"
                     /function="regulation raffinose-operon"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="regulatory protein"
                     /protein_id="AAA25562.1"
                     /db_xref="GI:475107"
                     /translation="MNGEYKTLANKSFESNVLFFGQEACLPNYTYKGNNVRDSYVIHY
                     IQEGKGTFAAANHPATVLKAGDIFILPKGTPCFYQADNDQPWKYFWIGFSAGIRIEAM
                     LSGSLLAQKCYLRQVQNGHIYADLSELYKVLHIPNSLINDVLLGSLIYRLFYDLLRWY
                     PADATNIKVKSTEQFNLAVSYLQENYSTGCTIMDLCHYLNLSRSYLYTLFKTHANTSP
                     QKLLTKLRLEDAKQRLSTSNNSVQSIANMVGYKDSFTFSKAFKRYSGASPSYYRKSIG
                     I"
     -10_signal      complement(3837. .3842)
                     /gene="rafR"
     -35_signal      complement(3859. .3864)
                     /gene="rafR"
     gene            3909. .5918
                     /gene="rafP"
```

```
-35_signal      3909. .3914
                /gene="rafP"
                /note="putative"
-10_signal      3932. .3937
                /gene="rafP"
                /note="putative"
CDS             3993. .5918
                /gene="rafP"
                /function="raffinose transport protein"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="permease"
                /protein_id="AAA25563.1"
                /db_xref="GI:475108"
                /translation="MQEEHNYKWVGGRLIYGFGAKGNDAFYSILSGYLIIFITSHLFD
                TGNKALDNRMVSLVTLIIMVLRIVELFIDPFIGNAIDRTKNSPGHFRPWVVVGGTVSS
                IILLLLFTNLGGLYAKNAMIYLVVFAILYITMDIFYSFKDVGFWSMLPSLTTDSRERE
                KTATFARLGSTIGGGLVGVLVMPAVIFFSAKATSTGDNRGWFIFALIICLIALISAWG
                VGLGTREVDSDIRKNKQDTVGVMEIFKALAKNDQLLWAALAYLFYGVGINILGSLEVY
                YFTYIMGKPKSFSILSIINIFLGLIATSLFPVLSKKFSRKGVFAGCLVFMLGGIAIFT
                IAGSNLWLVLLAATMFGFPQQMVFLVVLMVITDSVEYGQLKLGHRDESLALSVRPLID
                KFGGAISNGVVGQIAIISGMTTGATASSITAAGQLHFKLTMFAFPALMLLIAIGIFSK
                QIFLTEEKHAEIVAELERTWRTKFDNTTDQVAEKVVTSLDLATPIAGQVIPLAQVNDP
                TFAAGTLGDGFAIKPSDGRILAPFDATVRQVFTTRHAVGLVGDNGIVLLIHIGLGTVK
                LRGTGFISYVEEGQHVQQGDELLEFWDPTIKQAGLDDTVIMTVTNSTEFTMMDWLVKP
                GQAVKATDNILQLHTKA"
gene            5976. .8177
                /gene="agaR"
CDS             5976. .8177
                /gene="agaR"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="alpha-galactosidase"
                /protein_id="AAA25564.1"
                /db_xref="GI:475109"
                /translation="MSLITVDQANRVFHLHNQTLSYIFAVEQGGTLSHLYFGGHVDHY
                HGELRYPRVDRGFSGNLPGSTDRTFSRDTLPKEYSTAGEMDYHLPAAIVRHTDGANAL
                YLVYQGYRIEAGKPKLSGLPAAFVEDETEAETLTIVLVDQVSQVEFDLQYTIYRDRPV
                VTRSVQVCNQGDHAVNLEKVASMQIDFTDRQFETITLPGAHANERHPERGSINYGIQT
                FGSLRGTSSHQMNPFLALVDHTTTEFSGDAYGFNLVYSGNHAFELEKDQLDQLHLMVG
                INSYNFNWQLKAGATFQTPEVLMVYTNKGLNAMSQAYHHLIRERVVRSEFKNQERPIV
                VNNWEATFFDFNEAKLKPIVDEAKQLGIEMFVLDDGWFGHRDDDNSSLGDWQVDHRKF
                PQGLNHFVKYVHEQGLKFGIWLEPEMISYDSKLYQQHPDYLMQVPGRSPSPSRNQYIL
                DLGRQAVRNNIFDQLDQLLKSKQIDYIKWDMNRHLSDIYSVALPPERQGEVYHRYVLG
                LYELLERLTTAYPHILFEGCSGGGGRFDAGMAYYMPQIWASDNTDAVARLTIQYGTSL
                AYPISLATAHVSVSPNQQTGRETSMSTRSAVAASGVLGYELDLTQLSSADKQIVQKQV
                VQYKQIRPLIQFGEFYRLKSPITSNQAAWMFVSPQQDEAIVMVFNLTSYAQPSLTKTK
                LVGLNPKLNYQNIATKAIFGGDELMQLGFYDPVVYQDYTTKVYHFKAVTEN"
terminator      8265. .8289
                /standard_name="terminator 3"
                /note="putative"
repeat_unit     9403. .9426
                /note="DR2; putative"
                /rpt_type=direct
repeat_unit     10396. .10442
                /note="DR1; putative"
                /rpt_type=direct
repeat_unit     10397. .10443
                /note="IR2; putative"
                /rpt_type=inverted
terminator      10681. .10716
                /standard_name="terminator 2"
                /note="putative"
gene            complement(10727. .11593)
                /gene="scrK"
CDS             complement(10727. .11593)
                /gene="scrK"
                /function="fructose 6-phosphotransferase"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="fructokinase"
                /protein_id="AAA25565.1"
                /db_xref="GI:475110"
                /translation="MLLGAIEAGGTKFVCATGAENGQVSDRISIPTTTPVETMTAVDD
                YFTTHPVDAIGIGSFGPIGVNPHDPKYGYITTTPKPGWGDFDFLGHLKSQFNIPLYWT
                TDVNEAAYGESMIGIAKDVPNSIYMTIGTGVGAGVISQNHIFNGRTHTELGHMRLNRL
                PGDDFKSNCPYHDICLEGLAAGPAVGKRTGKAGKDIPVDDPVWPIITDYIAQACVNLT
                VAFAPDKIILNGGVMNQRQLFPMIREKFAAYLNGYEEVPPLDDYIVPAGLGNNSGIAG
                GLLLAQAALKNA"
gene            complement(11673. .13832)
                /gene="agaS"
CDS             complement(11673. .13832)
                /gene="agaS"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="alpha-galactosidase"
                /protein_id="AAA25566.1"
                /db_xref="GI:475111"
                /translation="MPVEYDPKTGLINLHNDQISYVIQILAHRYPVHRYFGRYFSKQP
                YFEPMPSGSHAFANDPTERFPYSVTSLPLEYSTIGSGDYRQPAYVIKDANNQLLPILE
                YTGFSVNDQPINSRQLPPTVSKHTPVTTLVIHLTDAVTKLQMDLNYTIFENQPLILRS
                TTLRHHGTTNLQVTALSSAQLDLPTDQYTALTLSGTHAHEANPSFNRLHPGLQTVRSL
                RGTSGPQHQPFMALAEPNTTELAGTVIGCALAWSGNFDSTVEVDQYQHSRLTIGLEPD
                TFEWQLKPNSSFQTPEAVLTWTNTGFNGMSQVFHDFSYQLMPSQTNIPSVLNTWETLT
                FAVSESKVQHLIEHAHQLGLQMLVLDDGWFVNRNGENGQLGDWFVDPIKFPNGLNPLA
                QQAHHHRMKFGLWVEPEMITTNSQLYQQHPDWVLQYVDRTPITARHQLVLDLSQAAVR
                DHLITTLTNLVQNNQLDYLKWDMNRHLTQVGSTHLPAAQQGELYHRYVCGLYDILTRL
                KRACPKLIIENCSAGGGRFDFGMLPYTNQTWISDLTDPVDRATIENGFSYLFPPRIFS
                NHITASPNAQNGRITPFETRLQLACIGQLGLELNPKQLAPSEQQLLRGALIKYQQLKS
                TFIKAHFYRLPTTRHVVAWLIVTADKKQAICCYLNGLNSRVKTQHPLPLHYLDAELAY
                SDSSGNRYTGHQLNTMGIPLKPTNADFTSQLIYLCQN"
gene            complement(13852. .15989)
                /gene="scrA"
CDS             complement(13852. .15807)
                /gene="scrA"
                /function="enzyme II of the PTS system, sucrose specific"
                /note="enzyme IIabc"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /protein_id="AAA25567.1"
                /db_xref="GI:475112"
                /translation="MNHQEVADRVLNAIGKNNIQAAAHCATRLRLVIKDESKIDQQAL
                DDDADVKGTFETNGQYQIIIGPGDVDKVYDALIVKTGLKEVTPDDIKAVAAAGQNKNP
                LMDFLKVLSDIFIPIVPALVAGGLLMALNNVLTAEHLFMAKSVVEVYPGLKGIAEMIN
                AMASAPFTFLPILLGFSATKRFGGNPYLGATMGMIMVLPSLVNGYSVATTMAAGKMVY
                WNVFGLHVAQAGYQGQVLPVLGVAFILATLEKFFHKHIKGAFDFTFTPMFAIVITGFL
                TFTIVGPVLRTVSDALTNGLVGLYNSTGWIGMGIFGLLYSAIVITGLHQTFPAIETQL
                LANVAKTGGSFIFPVASMANIGQGAATLAIFFATKSQKQKALTSSAGVSALLGITEPA
                IFGVNLKMKFPFVFAAIASGIASAFLGLFHVLSVAMGPASVIGFISIASKSIPAFMLS
                AVISFVVAFIPTFIYAKRTLGDDRDQVKSPAPTSTVINVNDEIISAPVTGASESLKQV
                NDQVFSAEIMGKGAAIVPSSDQVVAPADGVITVTYDSHHAYGIKTTAGAEILIHLGLD
                TVNLNGEHFTTNVQKGDTVHQGDLLGTFDIAALKAANYDPTVMLIVTNTANYANVERL
                KVTNVQAGEQLVALTAPAASSVAATTV"
-10_signal      complement(15872. .15877)
                /gene="scrA"
-35_signal      complement(15894. .15989)
                /gene="scrA"
-35_signal      15993. .15998
                /gene="scrB"
gene            15993. .17568
                /gene="scrB"
-10_signal      16017. .16022
```

Query Match 18.8%; Score 287.6; DB 1; Length 21838;
Best Local Similarity 58.8%; Pred. No. 1.9e-67;
Matches 539; Conservative 0; Mismatches 364; Indels 13; Gaps 2;

```
Qy      4 atggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcgt 63
          |||||   |||||||||  | |||||||||  |   |   | ||  ||   | |||||  |||||
Db  15324 ATGGCTAGTGCGCCGTTTACTTTCTTGCCAATCCTCTTAGGATTTTCAGCAACGAAGCGC 15265
```

```
Qy      64 ttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagctt 123
           ||||| |||||     |  ||||| ||  || || || |||   ||||| || ||     |
Db   15264 TTCGGTGGCAACCCCTATCTGGGTGCTACG-ATGGGCATGATCATGGTCTTACCATCACT 15206

Qy     124 ggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtccct 183
           || ||||||||   ||| ||  | |||||||| || ||| | |||       |||  | |
Db   15205 AGTTAACGGCTATAGCGTTGCAACGACCATGGCAGCCGGCAAGATGGTTTACTGGAACGT 15146

Qy     184 gtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtggt 243
           ||||| ||| | ||||| |||||||| || || |||   ||||| || ||  | |  ||
Db   15145 CTTTGGGTTACACGTTGCACAAGCCGGCTATCAAGGCCAAGTGCTCCCAGTCTTAGGCGT 15086

Qy     244 ttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcaga 303
             | |  ||||| || ||  | || || ||| | || || |   | || ||  |    ||
Db   15085 CGCCTTCATTCTAGCTACCCTTGAAAAATTCTTCCATAAACACATTAAAGGGGCATTTGA 15026

Qy     304 cttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgccat 363
           ||||  | | || ||  || | |  |  || | || || ||||| || || |    | |
Db   15025 CTTCACGTTTACCCCGATGTTTGCCATTGTGATTACTGGTTTCCTAACTTTTACAATCGT 14966

Qy     364 tggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatga 423
           ||| || |   | ||   |||| ||||||   |  | ||||| ||   ||  | ||| |
Db   14965 TGGTCCCGTCTTACGAACGGTGAGCGATGCATTAACTAACGGTTTAGTAGGCTTATATAA 14906

Qy     424 tttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcac 483
             |   ||      | ||         |||| ||| |  | || ||  |||| || || ||
Db   14905 CAGCACCGGCTGGATTGGTATGGGAATCTTTGGTTTATTATATTCTGCAATTGTTATTAC 14846

Qy     484 tggtctgcaccagtccttcccgccaattgagctggagctgtt----------taacca 531
           |||||| || ||  | |||||  |||| ||     || ||||             ||
Db   14845 TGGTCTCCATCAAACTTTCCCAGCAATCGAAACCCAGTTGTTGGCCAATGTTGCCAAAAC 14786

Qy     532 gggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatg 591
           |||||  ||||| |||||| ||    || || ||||| || || |  ||  |||||  ||
Db   14785 TGGTGGTTCCTTTATCTTCCCAGTCGCCTCGATGGCCAACATTGGGCAAGGTGCCGCCAC 14726

Qy     592 tttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcagg 651
           ||| ||  | |||||     | ||||| ||||||  || | |||  |   | |  | ||
Db   14725 TTTAGCTATTTTCTTTGCCACTAAGAGCCAAAAGCAAAAAGCCCTGACTTCTTCAGCCGG 14666

Qy     652 tgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcg 711
           |||||  ||  | || || |||||||| ||||| || ||||| || |||||     ||
Db   14665 GGTCTCAGCGTTACTCGGAATTACGGAACCTGCTATTTTCGGGGTCAACCTCAAGATGAA 14606

Qy     712 ctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctt 771
           |  || ||  || | |       | || |||  || | |   |||||  | | ||| ||
Db   14605 ATTTCCATTTGTCTTTGCAGCGATTGCCTCAGGAATTGCTTCAGCTTTCTTAGGACTATT 14546

Qy     772 taatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgc 831
             |  |   | | || ||| ||||  | |||    || | ||  |  |||| |||| | |
Db   14545 CCACGTTCTGTCCGTGGCGATGGGTCCCGCATCAGTCATCGGATTCATTTCGATTGCTTC 14486

Qy     832 tccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgc 891
                ||       ||| || |  |||| ||  |  | ||| || |||||||     |
Db   14485 AAAGTCGATTCCAGCATTCATGCTCAGTGCTGTCATCTCGTTCGTCGTCGCATTTATTCC 14426

Qy     892 agcgattgcttatggc 907
           | |  ||   |||| |
Db   14425 AACTTTTATCTATGCC 14410


RESULT   6
PPSURFOP/c
LOCUS       PPSURFOP               21839 bp    DNA     linear   BCT 24-MAY-1994
DEFINITION  P.pentosaceus (PPE1.0) sucrose and raffinose operons.
ACCESSION   Z32771
VERSION     Z32771.1  GI:493728
KEYWORDS    alpha-galactosidase; alpha-glucosidase; enzyme IIabc; fructokinase;
            insertion element; permease; regulator; sucrase.
SOURCE      Pediococcus pentosaceus.
  ORGANISM  Pediococcus pentosaceus
            Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
            Pediococcus.
REFERENCE   1  (bases 1 to 21839)
  AUTHORS   Leenhouts,K.K.J., Bolhuis,A.A., Kok,J.J. and Venema,G.G.
  TITLE     The sucrose and raffinose operons of Pediococcus pentosaceus PPE1.0
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 21839)
  AUTHORS   Leenhouts,K.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1994) Leenhouts K. K., Institute for Biological
            Sciences, Department of Genetics, Kerklaan 30, Haren, The
            Netherlands, 9751 NN
COMMENT     On May 26, 1994 this sequence version replaced gi:475962.
FEATURES             Location/Qualifiers
     source          1. .21839
                     /organism="Pediococcus pentosaceus"
                     /strain="PPE1.0"
                     /db_xref="taxon:1255"
     repeat_unit     487. .510
                     /note="DR2"
                     /citation=[1]
                     /rpt_type=DIRECT
     repeat_unit     488. .561
                     /note="IR1"
                     /citation=[1]
                     /rpt_type=INVERTED
     repeat_unit     513. .561
                     /note="DR1"
                     /citation=[1]
                     /rpt_type=DIRECT
     terminator      2724. .2760
                     /standard_name="terminator 4"
                     /citation=[1]
     gene            complement(2929. .3864)
                     /gene="rafR"
     CDS             complement(2929. .3762)
                     /gene="rafR"
                     /standard_name="regulator"
                     /function="regulation raffinose-operon"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="regulator"
                     /protein_id="CAA83663.1"
                     /db_xref="GI:475963"
                     /db_xref="SWISS-PROT:P43465"
                     /translation="MNGEYKTLANKSFESNVLFFGQEACLPNYTYKGNNVRDSYVIHY
                     IQEGKGTFAAANHPATVLKAGDIFILPKGTPCFYQADNDQPWKYFWIGFSAGIRIEAM
                     LSGSLLAQKCYLRQVQNGHIYADLSELYKVLHIPNSLINDVLLGSLIYRLFYDLLRWY
                     PADATNIKVKSTEQFNLAVSYLQENYSTGCTIMDLCHYLNLSRSYLYTLFKTHANTSP
                     QKLLTKLRLEDAKQRLSTSNNSVQSIANMVGYKDSFTFSKAFKRYSGASPSYYRKSIG
                     I"
     -10_signal      complement(3837. .3842)
                     /gene="rafR"
                     /citation=[1]
     -35_signal      complement(3859. .3864)
                     /gene="rafR"
                     /citation=[1]
     gene            3909. .5918
                     /gene="rafP"
     -35_signal      3909. .3914
                     /gene="rafP"
                     /citation=[1]
     -10_signal      3932. .3937
                     /gene="rafP"
                     /citation=[1]
     CDS             3993. .5918
                     /gene="rafP"
                     /standard_name="permease"
                     /function="raffinose transport protein"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="permease"
                     /protein_id="CAA83664.1"
```

```
                /db_xref="GI:475964"
                /db_xref="SWISS-PROT:P43466"
                /translation="MQEEHNYKWVGGRLIYGFGAKGNDAFYSILSGYLIIFITSHLFD
                TGNKALDNRMVSLVTLIIMVLRIVELFIDPFIGNAIDRTKNSPGHFRPWVVVGGTVSS
                IILLLLFTNLGGLYAKNAMIYLVVFAILYITMDIFYSFKDVGFWSMLPSLTTDSRERE
                KTATFARLGSTIGGGLVGVLVMPAVIFFSAKATSTGDNRGWFIFALIICLIALISAWG
                VGLGTREVDSDIRKNKQDTVGVMEIFKALAKNDQLLWAALAYLFYGVGINILGSLEVY
                YFTYIMGKPKSFSILSIINIFLGLIATSLFPVLSKKFSRKGVFAGCLVFMLGGIAIFT
                IAGSNLWLVLLAATMFGFPQQMVFLVVLMVITDSVEYGQLKLGHRDESLALSVRPLID
                KFGGAISNGVVGQIAIISGMTTGATASSITAAGQLHFKLTMFAFPALMLLIAIGIFSK
                QIFLTEEKHAEIVAELERTWRTKFDNTTDQVAEKVVTSLDLATPIAGQVIPLAQVNDP
                TFAAGTLGDGFAIKPSDGRILAPFDATVRQVFTTRHAVGLVGDNGIVLLIHIGLGTVK
                LRGTGFISYVEEGQHVQQGDELLEFWDPTIKQAGLDDTVIMTVTNSTEFTMMDWLVKP
                GQAVKATDNILQLHTKA"
gene            5976. .8177
                /gene="agaR"
CDS             5976. .8177
                /gene="agaR"
                /standard_name="alpha-galactosidase"
                /citation=[1]
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="alpha-galactosidase"
                /protein_id="CAA83665.1"
                /db_xref="GI:475965"
                /db_xref="SWISS-PROT:P43467"
                /translation="MSLITVDQANRVFHLHNQTLSYIFAVEQGGTLSHLYFGGHVDHY
                HGELRYPRVDRGFSGNLPGSTDRTFSRDTLPKEYSTAGEMDYHLPAAIVRHTDGANAL
                YLVYQGYRIEAGKPKLSGLPAAFVEDETEAETLTIVLVDQVSQVEFDLQYTIYRDRPV
                VTRSVQVCNQGDHAVNLEKVASMQIDFTDRQFETITLPGAHANERHPERGSINYGIQT
                FGSLRGTSSHQMNPFLALVDHTTTEFSGDAYGFNLVYSGNHAFELEKDQLDQLHLMVG
                INSYNFNWQLKAGATFQTPEVLMVYTNKGLNAMSQAYHHLIRERVVRSEFKNQERPIV
                VNNWEATFFDFNEAKLKPIVDEAKQLGIEMFVLDDGWFGHRDDDNSSLGDWQVDHRKF
                PQGLNHFVKYVHEQGLKFGIWLEPEMISYDSKLYQQHPDYLMQVPGRSPSPSRNQYIL
                DLGRQAVRNNIFDQLDQLLKSKQIDYIKWDMNRHLSDIYSVALPPERQGEVYHRYVLG
                LYELLERLTTAYPHILFEGCSGGGGRFDAGMAYYMPQIWASDNTDAVARLTIQYGTSL
                AYPISLATAHVSVSPNQQTGRETSMSTRSAVAASGVLGYELDLTQLSSADKQIVQKQV
                VQYKQIRPLIQFGEFYRLKSPITSNQAAWMFVSPQQDEAIVMVFNLTSYAQPSLTKTK
                LVGLNPKLNYQNIATKAIFGGDELMQLGFYDPVVYQDYTTKVYHFKAVTEN"
terminator      8265. .8289
                /standard_name="terminator 3"
                /citation=[1]
repeat_unit     9403. .9426
                /note="DR2"
                /citation=[1]
                /rpt_type=DIRECT
gene            complement(9519. .10424)
                /gene="iso-IS30"
CDS             complement(9519. .10424)
                /partial
                /gene="iso-IS30"
                /standard_name="IS-element"
                /function="IS-30 (E.coli) like IS element"
                /citation=[1]
                /codon_start=1
                /transl_table=11
                /product="insertion element"
                /protein_id="CAA83666.1"
                /db_xref="GI:475966"
                /db_xref="SPTREMBL:Q52261"
                /translation="SDAVRTKRSASLKWCLPQTHLLGADLLSSITYSERIKIETFCEL
                GLSNIQMGVRLNRSPSTISYELSRCQPYQAELAQTDAEYKRSQCGRKTKLSDELKQKI
                LNHLRLSWSPGMIAHEFKLATKSIYNWLNQGRIGFSLNDLPEHGVRQRRNVDQRSKYN
                QSLGRSIEQRPMMFNQRNRIGDFELDTVVGPRGHSKAVLLTLIDRKSRFLWAYRLKDR
                TTATVNEALTKFLTTFNGPVHSFTVDRGTEFSGLVSLESQYGIKTYYCHAYTPADVVV
                MNALIGIYVVFILKGLVLSTLVLKI"
repeat_unit     10396. .10442
                /note="DR1"
                /citation=[1]
                /rpt_type=DIRECT
repeat_unit     10397. .10443
                /note="IR2"
                /citation=[1]
                /rpt_type=INVERTED
terminator      10681. .10716
                /standard_name="terminator 2"
                /citation=[1]
gene            complement(10727. .11593)
                /gene="scrK"
CDS             complement(10727. .11593)
                /gene="scrK"
                /standard_name="fructokinase"
                /function="fructose 6-phosphotransferase"
                /citation=[1]
                /codon_start=1
                /transl_table=11
                /product="fructokinase"
                /protein_id="CAA83667.1"
                /db_xref="GI:475967"
                /db_xref="SWISS-PROT:P43468"
                /translation="MLLGAIEAGGTKFVCATGAENGQVSDRISIPTTTPVETMTAVDD
                YFTTHPVDAIGIGSFGPIGVNPHDPKYGYITTTPKPGWGDFDFLGHLKSQFNIPLYWT
                TDVNEAAYGESMIGIAKDVPNSIYMTIGTGVGAGVISQNHIFNGRTHTELGHMRLNRL
                PGDDFKSNCPYHDICLEGLAAGPAVGKRTGKAGKDIPVDDPVWPIITDYIAQACVNLT
                VAFAPDKIILNGGVMNQRQLFPMIREKFAAYLNGYEEVPPLDDYIVPAGLGNNSGIAG
                GLLLAQAALKNA"
gene            complement(13852. .15989)
                /gene="scrA"
CDS             complement(13852. .15807)
                /gene="scrA"
                /standard_name="enzyme IIscr"
                /function="enzyme II of the PTS system, sucrose specific"
                /citation=[1]
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="enzyme IIabc"
                /protein_id="CAA83668.1"
                /db_xref="GI:475968"
                /db_xref="SWISS-PROT:P43470"
                /translation="MNHQEVADRVLNAIGKNNIQAAAHCATRLRLVIKDESKIDQQAL
                DDDADVKGTFETNGQYQIIIGPGDVDKVYDALIVKTGLKEVTPDDIKAVAAAGQNKNP
                LMDFLKVLSDIFIPIVPALVAGGLLMALNNVLTAEHLFMAKSVVEVYPGLKGIAEMIN
                AMASAPFTFLPILLGFSATKRFGGNPYLGATMGMIMVLPSLVNGYSVATTMAAGKMVY
                WNVFGLHVAQAGYQGQVLPVLGVAFILATLEKFFHKHIKGAFDFTFTPMFAIVITGFL
                TFTIVGPVLRTVSDALTNGLVGLYNSTGWIGMGIFGLLLYSAIVITGLHQTFPAIETQL
```

```
  Query Match             18.8%;  Score 287.6;  DB 1;  Length 21839;
  Best Local Similarity   58.8%;  Pred. No. 1.9e-67;
  Matches  539;  Conservative    0;  Mismatches  364;  Indels   13;  Gaps    2;

Qy       4 atggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcgt 63
           |||||   |||||||||  | |||||||||| |  |  | || ||  | ||||| |||||
Db   15324 ATGGCTAGTGCGCCGTTTACTTTCTTGCCAATCCTCTTAGGATTTTCAGCAACGAAGCGC 15265

Qy      64 ttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagctt 123
           |||||  |||||    |  |||||  ||  || ||  ||  |||   |||||  ||  ||     |
Db   15264 TTCGGTGGCAACCCCTATCTGGGTGCTACG-ATGGGCATGATCATGGTCTTACCATCACT 15206

Qy     124 ggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtccct 183
           ||  ||||||||   |||  ||  | ||||||||  ||  |||  | |||      |||  | |
Db   15205 AGTTAACGGCTATAGCGTTGCAACGACCATGGCAGCCGGCAAGATGGTTTACTGGAACGT 15146

Qy     184 gtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtggt 243
           |||||  |||  |  |||||  ||||||||  ||  ||  |||   |||||  ||  ||  |  |  ||
Db   15145 CTTTGGGTTACACGTTGCACAAGCCGGCTATCAAGGCCAAGTGCTCCCAGTCTTAGGCGT 15086

Qy     244 ttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcaga 303
             |  |  |||||  ||  ||   |  ||  ||  |||  |  ||  ||  |    |  ||  ||   |     ||
Db   15085 CGCCTTCATTCTAGCTACCCTTGAAAAATTCTTCCATAAACACATTAAAGGGGCATTTGA 15026

Qy     304 cttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgccat 363
           ||||  |  |  ||  ||  ||  |  |  |  ||  |  |  ||  ||  |||||  ||  ||  |    |  |
Db   15025 CTTCACGTTTACCCCGATGTTTGCCATTGTGATTACTGGTTTCCTAACTTTTACAATCGT 14966
```

```
Qy     364 tggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatga 423
           ||| || |   | ||   |||| ||||||   |  |  ||||| ||   ||  | ||| |
Db   14965 TGGTCCCGTCTTACGAACGGTGAGCGATGCATTAACTAACGGTTTAGTAGGCTTATATAA 14906

Qy     424 tttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcac 483
           |   ||     | ||        |||| ||| |  | || ||  |||| || || ||
Db   14905 CAGCACCGGCTGGATTGGTATGGGAATCTTTGGTTTATTATATTCTGCAATTGTTATTAC 14846

Qy     484 tggtctgcaccagtccttcccgccaattgagctggagctgtt-----------taacca 531
           |||||| || ||  | |||||  |||| ||     || ||||             ||
Db   14845 TGGTCTCCATCAAACTTTCCCAGCAATCGAAACCCAGTTGTTGGCCAATGTTGCCAAAAC 14786

Qy     532 gggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatg 591
           ||||| ||||| |||||| ||   || || ||||| || || |  || ||||| ||
Db   14785 TGGTGGTTCCTTTATCTTCCCAGTCGCCTCGATGGCCAACATTGGGCAAGGTGCCGCCAC 14726

Qy     592 tttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcagg 651
           ||| ||  | |||||     | |||||  ||||||  || | |||  |    | |  | ||
Db   14725 TTTAGCTATTTTCTTTGCCACTAAGAGCCAAAAGCAAAAAGCCCTGACTTCTTCAGCCGG 14666

Qy     652 tgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcg 711
           |||||  ||  | || || |||||||| ||||| || ||||| || |||||     ||
Db   14665 GGTCTCAGCGTTACTCGGAATTACGGAACCTGCTATTTTCGGGGTCAACCTCAAGATGAA 14606

Qy     712 ctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctt 771
           |  || ||  || | |       | || |||  || | |   |||||  | | ||| ||
Db   14605 ATTTCCATTTGTCTTTGCAGCGATTGCCTCAGGAATTGCTTCAGCTTTCTTAGGACTATT 14546

Qy     772 taatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgc 831
           |  |   | | || ||| ||||  | |||    || | ||  |  |||| |||| | |
Db   14545 CCACGTTCTGTCCGTGGCGATGGGTCCCGCATCAGTCATCGGATTCATTTCGATTGCTTC 14486

Qy     832 tccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgc 891
           |         ||| || |  |||| ||  |  | ||| || |||||||     |
Db   14485 AAAGTCGATTCCAGCATTCATGCTCAGTGCTGTCATCTCGTTCGTCGTCGCATTTATTCC 14426

Qy     892 agcgattgcttatggc 907
           | |  ||   |||| |
Db   14425 AACTTTTATCTATGCC 14410
```

```
RESULT   7
AF401046
LOCUS       AF401046                2788 bp    DNA     linear   BCT 21-AUG-2001
DEFINITION  Lactobacillus sakei sucrose-specific enzyme II of the PTS (scrA)
            and dextran glucosidase (dexB) genes, partial cds.
ACCESSION   AF401046
VERSION     AF401046.1  GI:15217137
KEYWORDS    .
SOURCE      Lactobacillus sakei.
  ORGANISM  Lactobacillus sakei
            Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
            Lactobacillus.
REFERENCE   1  (bases 1 to 2788)
  AUTHORS   Dudez,A.-M., Chaillou,S., Hissler,L., Stentz,R.,
            Champomier-Verges,M., Alpert,C.-A. and Zagorec,M.
  TITLE     Physical and genetic map of the Lactobacillus sakei 23K chromosome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2788)
  AUTHORS   Dudez,A.-M., Chaillou,S. and Zagorec,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Flore Lactique et Environnement Carne,
            INRA, Domaine de Vilvert, Jouy-en-Josas 78350, France
FEATURES             Location/Qualifiers
     source          1. .2788
                     /organism="Lactobacillus sakei"
                     /db_xref="taxon:1599"
     gene            <1. .1605
                     /gene="scrA"
     CDS             <1. .1605
                     /gene="scrA"
                     /note="ScrA; similar to ScrA of Pediococcus pentosaceus"
                     /codon_start=1
                     /transl_table=11
                     /product="sucrose-specific enzyme II of the PTS"
                     /protein_id="AAK92528.1"
                     /db_xref="GI:15217138"
                     /translation="FVPLIPALTAGGLLMAINNVLTGQGLFGAQSIVQMFPQWKGFAE
                     IVNMMSSAPFTFLPILIAFSATKRFGGNPYLGAAAGMMLVMPNLVNGYGVAESIATGH
                     MTYWHVFGLNIAQAGYQGQVIPVIGVAFILANLEKFFHKHLNDAVDFTFTPMLSIIIT
                     GFLTFTLVGPALRIVSNGVTDSLVWAYQTLGAVGMGIFGLGYSAIVLTGLHQSFPAIE
                     TTLLADIAKTGGSFIFPVAAMANIAQGAATFAVFFVTKNKQQKSLTTSAGISAMLGIT
                     EPALFGVNLKLKFPFFIGLIASGISSFIIGLLHVLSVSMGPAGIIGFIAIAPKSIPSF
                     MMGAIISFVIAFVGTYLYGKKAMKTTEEEIINEAPATPEVVERLQDEKISAPVTGRIV
                     DLASVPDPVFASEAMGKGIAIMPTSQDVLAPVTGVITIAANTGHAYGIKSDDGAEVLI
                     HIGLDTVNLNGIGFEKIVQQGQHVSEGDLLGHFDIDKIKQAGLTPLTMTIVTNTAGYA
                     QVDPLLTVDKAAMQGEEIIQLHAKKD"
     gene            1604. .>2788
                     /gene="dexB"
     RBS             1604. .1610
                     /gene="dexB"
     CDS             1617. .>2788
                     /gene="dexB"
                     /note="DexB; similar to DexB of Streptococcus mutans"
                     /codon_start=1
                     /transl_table=11
                     /product="dextran glucosidase"
                     /protein_id="AAK92529.1"
                     /db_xref="GI:15217139"
                     /translation="MQTNWWQNAVFYQVYPRSFQDSNGDGIGDIQGIIQRLDYLADLG
                     VNAIWLSPVYQSPNVDNGYDISDYQAINPEYGSMVDMEQLIEAAKIRKIKIVMDLVVN
                     HTSDQHPWFLEARKSKDNPYRDFYIWRDPATDGSVPNDLQSNFKGSAWAFDAVTGQYY
                     LHFYAKEQPDLNWQNPKVREAVYQMMTWWLQKGIGGFRMDVIDLIGKEPDRKIKENGP
                     QLHAYLQEMNARVLSQYDVVTVGETWGATPEIGQMYSNPNRHELSMIFQFEQINLDKQ
                     SGMTRWDLKPLIPAELHAVFSKWQLALDGVGWNSLFWSNHDLPRIVSRWGDDSQYREK
                     SAQALAIYLHMLKGTPYIYQGEEIGMTNYPITSYHEIEDIESRRVYQQRQQQGYAI"
BASE COUNT      842 a    451 c    637 g    858 t
ORIGIN
```

```
  Query Match           16.6%;  Score 253.4;  DB 1;  Length 2788;
  Best Local Similarity   51.9%;  Pred. No. 3.4e-58;
  Matches  734;  Conservative    0;  Mismatches  621;  Indels   58;  Gaps    5;

Qy       2 tcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagc 61
           | ||| | || ||||| ||  | ||| | ||| | ||  |||  ||  | ||||| ||||
Db     143 TGATGTCGTCAGCGCCATTTACCTTCCTACCAATTTTAATTGCATTTTCTGCAACTAAGC 202

Qy      62 gtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagc 121
           | || || || |||    |   | || || ||     || |||  | ||||  | || | |
Db     203 GATTTGGTGGAAATCCATATTTAGGTGCGGCAGCA-GGAATGATGTTGGTTATGCCTAAC 261

Qy     122 ttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtcc 181
           | || || || || |  || || |  |  || ||| | ||  | ||| |    |||
Db     262 CTAGTTAATGGTTATGGGGTTGCTGAAAGTATCGCTACAGGGCATATGACGTATTGGCAT 321

Qy     182 ctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtg 241
           | ||||| ||  || |||| ||||| ||||| || ||    ||  |||| ||  | |
Db     322 GTCTTTGGATTGAATATTGCGCAAGCTGGTTATCAAGGTCAAGTAATTCCGGTTATCGGA 381

Qy     242 gtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgca 301
           ||| | |  ||||| || |   | ||||| ||| | || || |   | || |  |||
Db     382 GTTGCCTTCATTCTTGCTAATTTAGAGAAATTCTTCCATAAACATTTAAATGATGCTGTT 441

Qy     302 gacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcc 361
           || ||   | | || ||| || |  || |  |  |||| |||||  | || || |
Db     442 GATTTTACGTTTACACCAATGTTATCGATTATTATCACAGGATTTTTAACCTTTACATTA 501

Qy     362 attggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttat 421
           |||| || ||| || |    ||     |||   |  |  |  || ||    |  ||||
Db     502 GTTGGTCCTGCATTGAGAATTGTTTCAAATGGTGTTACTGATAGTTTAGTTTGGGCTTAT 561

Qy     422 gatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatc 481
           |       ||| |||||||     | | || || ||||  || ||| |||| ||| |
Db     562 CAAACGTTAGGTGCAGTCGGTATGGGGATTTTTGGCCTGGGTTATTCAGCAATTGTCTTA 621
```

```
Qy    482 actggtctgcaccagtccttcccgccaattgagctggagctgttt----------aac 529
          |||||  | || ||   ||||||| |||||||   |   || ||           ||
Db    622 ACTGGCTTACATCAAAGCTTCCCGGCAATTGAAACGACACTTTTGGCAGATATTGCCAAA 681

Qy    530 cagggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggca 589
          ||||||||  || || ||  |    ||| | ||||| ||||| || || || || |||
Db    682 ACTGGTGGATCGTTTATTTTTCCCGTTGCAGCGATGGCAAATATTGCTCAAGGGGCTGCA 741

Qy    590 tgtttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttca 649
          |||| || || |||||| |  | |||| | || | |  |||    |  |   | || |
Db    742 ACTTTCGCTGTATTCTTCGTTACTAAGAATAAACAACAAAAGTCATTAACGACTTCTGCT 801

Qy    650 ggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctg 709
          ||  | || ||  |  | || ||||| || || ||  | || || || ||  |     ||
Db    802 GGGATTTCTGCGATGTTGGGAATTACTGAACCAGCATTATTTGGGGTTAATTTAAAATTG 861

Qy    710 cgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactc 769
          |  || ||||| || ||| |   | |  |||  |||         | | ||||   |
Db    862 AAGTTTCCATTCTTTATTGGTTTAATTGCATCAGGAATCTCATCGTTTATTATTGGTTTA 921

Qy    770 tttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgat 829
          ||  || |    ||||  |  ||||  |||||||  |  | ||  || || | ||||
Db    922 TTACATGTTTTATCAGTATCAATGGGACCTGCAGGAATTATTGGGTTTATTGCGATTGCA 981

Qy    830 gctccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggc 889
          ||     ||    |  ||  || || | ||  || |||  ||| | || || || | |
Db    982 CCTAAGAGCATCCCTAGTTTTATGATGGGAGCTATTATTAGTTTCGTAATTGCCTTTGTG 1041

Qy    890 gcagcgattgcttatggcctttacttggttcgccgcaacggcagcattgatccagatgca 949
          |   |       || ||                                 |  | | |||
Db   1042 GGGACATACTTATACGG--------------------------------TAAAAAGGCA 1068

Qy    950 accgctgctccagtgcctgcaggaacgaccaaagccgaagcagaagcacccgcagaattt 1009
          |     | | |    | |  ||  |   ||||  |||    ||   | ||
Db   1069 ATGAAGACAACTGAAGAAGAAATAATCAATGAAGCACCAGCTACCCCAGAAGTAGTGGAG 1128

Qy   1010 tcaaacgattccaccatcatccaggcacctttgaccggtgaagctattgcactgagcagc 1069
            |     |   |  ||    ||||| | |||||   |  | |||   |
Db   1129 AGATTACAAGATGAAAAGATTAGTGCACCAGTTACCGGACGAATTGTTGACTTAGCATCA 1188

Qy   1070 gtcagcgatgccatgtttgccagcggaaagcttggctcgggcgttgccatcgtcccaacc 1129
          ||    ||| |  | ||||| || | |    | ||    ||| |||| ||  | |||||
Db   1189 GTACCTGATCCAGTTTTTGCAAGTGAAGCAATGGGAAAAGGCATTGCGATTATGCCAACT 1248

Qy   1130 aaggggcagttagtttctccggtgagtggaaagattgtggtggcattcccatctggccat 1189
            | |  || || | || || | ||      |||| |  || |  | ||     |
Db   1249 TCTCAGGATGTACTTGCACCAGTTACCGG------TGTGATAACAATTGCGGCTAATACT 1302

Qy   1190 gctttcgcagttcgcaccaaggctgaggatggttccaatgtggatatcttgatgcacatt 1249
          | |  ||||    | |  ||  | || |||||| |  | |||       | || || |||
Db   1303 GGTCACGCATACGGGATAAAATCGGATGATGGTGCAGAAGTG------CTAATTCATATT 1356

Qy   1250 ggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcagggcgat 1309
          ||||| || ||||| ||  | || || |     ||| |   | |     | |||||  |
Db   1357 GGTTTAGATACAGTTAATTTAAATGGTATAGGTTTTGAAAAGATTGTCCAACAGGGACAA 1416

Qy   1310 gaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgcaggt 1369
          | || |  | ||| ||  |  |  || | || |||||||||   |||||    || ||
Db   1417 CATGTTAGCGAAGGCGATTTATTAGGTCATTTTGATATTGATAAGATTAAACAAGCCGGG 1476

Qy   1370 tatgaggtaaccacgccgattgttgtttcgaat 1402
                ||   || | ||||  |||||
Db   1477 CTAACACCGCTAACAATGACTATTGTGACGAAT 1509
```

```
RESULT   8
LLZ97015
LOCUS       LLZ97015               5800 bp    DNA     linear   BCT 05-APR-1999
DEFINITION  Lactococcus lactis cremoris sucrose gene cluster.
ACCESSION   Z97015
VERSION     Z97015.1  GI:4581474
KEYWORDS    enzyme II sucrose protein; fructokinase; sacA gene; sacB gene; sacK
            gene; sacR gene; sucrose-6-phosphate hydrolase.
SOURCE      Lactococcus lactis.
  ORGANISM  Lactococcus lactis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Lactococcus.
REFERENCE   1  (bases 1 to 5800)
  AUTHORS   Luesink,E.J., Marugg,J.D., Kuipers,O.P. and de Vos,W.M.
  TITLE     Characterization of the divergent sacBK and sacAR operons, involved
            in sucrose utilization by Lactococcus lactis
  JOURNAL   J. Bacteriol. 181 (6), 1924-1926 (1999)
  MEDLINE   99173919
REFERENCE   2  (bases 1 to 5800)
  AUTHORS   Luesink,E.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1997) Luesink E.J., Department of Biophysical
            Chemistry, Netherlands Institute for Dairy Research (NIZO), P.O.
            Box 20, Ede 6714 BA, THE NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .5800
                     /organism="Lactococcus lactis"
                     /sub_species="cremoris"
                     /db_xref="taxon:1358"
     gene            complement(330. .1286)
                     /gene="sacR"
     CDS             complement(330. .1286)
                     /gene="sacR"
                     /function="negative regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="SacR"
                     /protein_id="CAB09692.1"
                     /db_xref="GI:4581475"
                     /translation="MIKLEDVANKAGVSVTTVSRVINRKGYLSDATISKVEKAMQDLH
                     YIPNAAARSLQGKSLKLIGLVFPTIKNIFYAELIEKIEQALFIRGYKAMLATTEHDEQ
                     KERDYLALLLSNQVDGIIYGSHNLKAHDYIAIEAPIVAFDRLLTPETTVVSSDNFEGG
                     ILATKALINSGSKKIAIFTGNDNTNSPTYLRRDGYLLELERNQLKPHIIKIPSQWTLL
                     RKKVEIKKILENNDFDGVFCTDDLTAILVKDLASNLKKSLNVVGFDGTEFIENYYPNL
                     TTIKQPINDLAELLVDLIIRKIDGDNIDITYQLPVQLHYGID"
     gene            complement(join(1283. .2704,2721. .2725,2737. .2743,
                     2761. .2767))
                     /gene="sacA"
     CDS             complement(1283. .2704)
                     /gene="sacA"
                     /EC_number="3.2.1.26"
                     /codon_start=1
                     /transl_table=11
                     /product="sucrose-6-phosphate hydrolase"
                     /protein_id="CAB09689.1"
                     /db_xref="GI:4581476"
                     /translation="MKWSTKQRYRTYDSYSESDLESLRKLALKSPWKSNFHIEPETGL
                     LNDPNGFSYFNEKWHLFYQHFPFGPVHGLKSWVHLVSDDLVHFEKTGLVLYPDTKYDN
                     AGVYSGSALAFENFLFLIYTGNHRGEDWVRTPYQLGAKIDKNNQLVKFTEPLIYPDFS
                     QTTDHFRDPQIFSFQGQIYCLIGAQSSQKNGIIKLYKAIENNLTDWKDLGNLDFSKEK
                     MGYMIECPNLIFINGRSVLVFCPQGLDKSIVKYDNIYPNVYVIADDFTTGSKNQLKNA
                     GQLINLDEGFDCYATQSFNAPDGSAYAISWLGLPETSYPTDKYNVQGVLSMVKKLSIK
                     DNKLYQYPVEKMKELRQMEQDLLLADNNIITSNSYELEVDFRQQTSTLLSLMTNEKGD
                     SALKVEIDKENNTITLIRNYEKRLAHVKIEKMNVFIDQSIFEIFINDGEKVLSDCRVF
                     PNKNQYSIRSQNPIKIKLWELKK"
     RBS             complement(1294. .1298)
                     /gene="sacR"
     gene            complement(1294. .1298)
                     /gene="sacR"
     gene            complement(1390. .1396)
                     /gene="sacR"
     -10_signal      complement(1390. .1396)
                     /gene="sacR"
                     /evidence=experimental
     -35_signal      complement(1416. .1422)
                     /gene="sacR"
                     /evidence=experimental
     gene            complement(1416. .1422)
                     /gene="sacR"
     RBS             complement(2721. .2725)
                     /gene="sacA"
```

```
     -10_signal      complement(2737. .2743)
                     /gene="sacA"
                     /evidence=experimental
     -35_signal      complement(2761. .2767)
                     /gene="sacA"
                     /evidence=experimental
     -35_signal      2815. .2821
                     /gene="sacB"
                     /evidence=experimental
     gene            2815. .4835
                     /gene="sacB"
     -10_signal      2837. .2843
                     /gene="sacB"
                     /evidence=experimental
     RBS             2872. .2877
                     /gene="sacB"
     CDS             2883. .4835
                     /gene="sacB"
                     /EC_number="2.7.1.69"
                     /codon_start=1
                     /transl_table=11
                     /product="enzyme II sucrose protein"
                     /protein_id="CAB09690.1"
                     /db_xref="GI:4581477"
                     /translation="MNHKQVAERILNAVGRDNIQGARHCATRLRLVLKDTGVIDQEAL
                     DNDPDLKGTFEAAGQYQIIVGPGDVNTVYEEFIKLTGISEASTADLKEIAGSQKKQNP
                     VMALVKLLSDIFVPLIPALVAGGLLMALNNALTAEHLFATKSLVEMFPMWKGFADIVN
                     TMSAAPFTFMPILIGYSATKRFGGNPYLGAVVGMIMVMPGLINGYNVAEAISNHTMTY
                     WDIFGFKVAQAGYQGQVLPVIGVAFILAKLERFFHKYLNDAIDFTFTPLLSVIITGFL
                     TFTIVGPALRFVSNGLTDGLVGLYNTLGALGMLVFGGFYSAIVVTGLHQSFPAIETML
                     ITNYQHSGIGGDFIFPVAACANMAQAGATFAILFVTKNIKTKALAAPAGVSAILGITE
                     PALFGINLKLKYPFFIALGASAIGSLFMGLFHVLAVSLGSAGLIGFISIKAGYNLQFM
                     ISIFISFLIAFVVTSIYGRRMEAKSITKEKNKQNATTQYQPEKVIIDPVKSGELLAPI
                     NGFVIPLSDVSDPVFSKEIMGKGIAIKPKSGELFSPADGEIIIAYETGHAYGIKTKNG
                     GEVLLHIGIDTVSMNGNGFIQNVKVGQKVKAGDLLGSFDKEEIKKSGLDDTVIIVITN
                     SASYNEILPLSENVDIKVGEKILLLN"
     RBS             4838. .4842
                     /gene="sacK"
     gene            4838. .5720
                     /gene="sacK"
     CDS             4848. .5720
                     /gene="sacK"
                     /EC_number="2.7.1.4"
                     /codon_start=1
                     /transl_table=11
                     /product="fructokinase"
                     /protein_id="CAB09691.1"
                     /db_xref="GI:4581478"
                     /translation="MSVYYGSIEAGGTKFVLAIADEHFNIIKKFKFATTTPQETISKT
                     IKYFKENRVSAIGLGSFGPIDLNLSSKTYGYITSTPKVGWKNINLVGQLKEALDIPIY
                     FTTDVNASAYGEMKNTGIKNLVYLTIGTGIGGGAIQNGYFIGGIGHSEMGHQRINRHR
                     DVLTFEGICPFHGDCLEGVAAGPSLEARTGILGEKISSDDPIWDILSYYIAQAAINAT
                     LTLAPECIILGGGVMEKPNMISLIQKQFISMLNNYIDLPCSVEKYIRLPTVKENGSAT
                     LGNFYLAYSLFTKE"
BASE COUNT     1858 a    915 c   1031 g   1996 t
ORIGIN
```

```
  Query Match            16.4%;  Score 250.8;  DB 1;  Length 5800;
  Best Local Similarity  51.1%;  Pred. No. 2e-57;
  Matches  725;  Conservative    0;  Mismatches  647;  Indels   46;  Gaps    4;

Qy       4 atggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcgt 63
           |||    | || || ||  | ||  ||||| | ||  | ||||    | || || ||  |
Db    3366 ATGAGTGCCGCACCATTTACCTTTATGCCAATTTTAATAGGTTATTCTGCCACAAAAAGA 3425

Qy      64 ttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagctt 123
           || ||||| |||   |   | || || | | | | ||||||   |||||| | ||  |  |
Db    3426 TTTGGCGGAAATCCATATTTAGGTGCTG-TTGTAGGTATGATTATGGTGATGCCTGGACT 3484

Qy     124 ggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtccct 183
             | || || ||  | || || |   | ||  ||         |||| |    |||    |
Db    3485 TATTAATGGATATAATGTTGCTGAAGCAATCTCTAATCATACAATGACTTATTGGGATAT 3544
```

```
Qy     184 gtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtggt 243
            |||||||| | |||||||| || || || || ||    || |||||||  | |  ||
Db    3545 TTTTGGTTTTAAAGTTGCCCAGGCTGGATATCAAGGACAAGTTCTTCCTGTAATTGGAGT 3604

Qy     244 ttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcaga 303
            |||  || || || |   | ||||  ||| | || ||    || || |   ||  |||
Db    3605 GGCTTTTATCCTTGCTAAACTTGAGAGATTCTTTCATAAATACCTTAACGATGCTATAGA 3664

Qy     304 cttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgccat 363
            ||||    | |||||  | ||  |  |  |  | || |||||  || |||||||      |
Db    3665 CTTCACATTTACTCCGTTACTTTCAGTCATCATAACTGGATTTCTCACATTCACTATTGT 3724

Qy     364 tggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatga 423
            || |||||| | || |  ||     |||   || || | ||| |    |||||||| |
Db    3725 GGGACCAGCACTTCGTTTTGTATCAAATGGCTTGACAGATGGTTTGGTTGGACTTTATAA 3784

Qy     424 tttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcac 483
               || || | ||     ||  | || ||      |||||||  |||| ||  | ||
Db    3785 CACATTAGGGGCACTAGGAATGCTTGTGTTTGGAGGCTTCTACTCGGCAATAGTAGTAAC 3844

Qy     484 tggtctgcaccagtccttcccgccaattgagctggagctgtttaaccag---------- 532
            |||  | || |||   || ||| | |||||      | |  | |   |
Db    3845 TGGATTACATCAGAGTTTTCCGGCCATTGAAACAATGTTAATCACAAACTATCAACACAG 3904

Qy     533 -------ggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgc 585
                  ||||||  ||| || ||  ||   ||  |   || ||||| || || | ||
Db    3905 TGGTATTGGTGGAGACTTTATTTTTCCAGTTGCGGCCTGCGCAAATATGGCTCAAGCTGG 3964

Qy     586 ggcatgtttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgc 645
            |||  ||| ||| | || ||  |  | || |     |||   || |  |||||||  | |
Db    3965 TGCAACTTTTGCAATTTTATTTGTTACTAAAAACATTAAGACAAAAGCTCTTGCAGCTCC 4024

Qy     646 ttcaggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcg 705
            | ||||| || ||| |||| |||||||| || |||||| | || ||  | || ||
Db    4025 AGCTGGTGTATCTGCTATTCTAGGTATTACAGAACCTGCGTTATTTGGGATTAATCTAAA 4084

Qy     706 cctgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgc 765
            ||    | |||||||| || | | | ||  |  |||| || |||       || || |
Db    4085 ACTAAAATATCCGTTCTTTATTGCTCTTGGGGCTTCAGCAATTGGTTCATTATTTATGGG 4144

Qy     766 actctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctat 825
            | | ||  || ||  || |||    ||||  | ||||| ||  | || ||  | |||||
Db    4145 ATTATTCCATGTCCTTGCGGTTAGTCTGGGATCCGCAGGATTAATTGGCTTTATCTCTAT 4204

Qy     826 tgatgctccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcatt 885
            | | |||    |||  |    ||  || | |    | || ||| |||  | || || ||
Db    4205 TAAAGCTGGGTATAACTTACAATTTATGATTTCGATATTTATTAGCTTTCTTATTGCCTT 4264

Qy     886 cggcgcagcgattgcttatggcctttacttggttcgccgcaacggcagcattgatccaga 945
            |  |   |       ||||| |      |||                |  |  ||   | |
Db    4265 TGTTGTTACCTCAATCTATGGTCGGCGGATGGAAGCTAAATCTATTACGAAAGAAAAAAA 4324

Qy     946 tgcaaccgctgctccagtgcctgcaggaacgaccaaagccgaagcagaagcacccgcaga 1005
            |  |     |||                |||  ||  ||  | || | ||   |
Db    4325 TAAACAGAATGC--------------AACAACTCAATACCAACCTGAGAAAGTTATTAT 4369

Qy    1006 attttcaaacgattccaccatcatccaggcacctttgaccggtgaagctattgcactgag 1065
            | ||    |       |   ||| || ||  |  ||     |  ||| | |||
Db    4370 CGATCCAGTTAAAAGTGGCGAACTCCTTGCTCCGATAAATGGATTTGTGATTCCTCTGTC 4429

Qy    1066 cagcgtcagcgatgccatgtttgccagcggaaagcttggctcgggcgttgccatcgtccc 1125
               || || ||  |  | ||  | |  | ||   | ||    ||  |||| |||   ||
Db    4430 TGATGTAAGTGACCCTGTTTTCTCAAAAGAAATTATGGGAAAAGGTATTGCAATCAAGCC 4489

Qy    1126 aaccaaggggcagttagtttctccggtgagtggaaagattgtggtggcattcccatctgg 1185
            |      ||  |  |  | || || |      ||| | ||| |  |  || |    | | ||
Db    4490 TAAATCTGGAGAACTTTTCTCACCAGCCGACGGAGAAATTATTATTGCTTATGAAACAGG 4549

Qy    1186 ccatgctttcgcagttcgcaccaaggctgaggatggttccaatgtggatatcttgatgca 1245
            || |||| |   |    || ||  ||  | |          ||  |  |  | ||
Db    4550 TCACGCTTATGGTATAAAAACAAAAAATGGAGGT----------GAAGTTCTTTTACA 4597
```

```
Qy    1246 cattggtttcgacacagtaaacctcaacggcacgcactttaacccgctgaagaagcaggg 1305
           |||||  | || || ||     | || || |     || |  |        ||    ||
Db    4598 TATTGGAATAGATACTGTTTCAATGAATGGTAATGGATTCATACAAAATGTTAAAGTTGG 4657

Qy    1306 cgatgaagtcaaagcaggggagctgctgtgtgaattcgatattgatgccattaaggctgc 1365
           | |  |||| ||||||||||| | ||  |    || ||||  || |  || |||
Db    4658 CCAGAAAGTAAAAGCAGGGGATTTACTAGGATCTTTTGATAAAGAAGAAATCAAGAAGAG 4717

Qy    1366 aggttatgaggtaaccacgccgattgttgtttcgaatt 1403
           || |  || |  ||       ||||| || | ||||
Db    4718 TGGATTGGATGATACTGTAATTATTGTTATTACAAATT 4755


RESULT  9
AE004395/c
LOCUS       AE004395               10264 bp    DNA     linear   BCT 31-JUL-2000
DEFINITION  Vibrio cholerae chromosome II, section 52 of 93 of the complete
            chromosome.
ACCESSION   AE004395 AE003853
VERSION     AE004395.1  GI:9658068
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 10264)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 10264)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10264
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="II"
                     /note="biotype: El Tor"
     gene            complement(114. .1553)
                     /gene="VCA0653"
     CDS             complement(114. .1553)
                     /gene="VCA0653"
                     /note="similar to GB:M76768 SP:P22825 PID:155262;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="PTS system, sucrose-specific IIBC component"
                     /protein_id="AAF96554.1"
                     /db_xref="GI:9658069"
                     /translation="MDYPVIAKQLLESLGGKSNIQALAHCATRLRLVLNDETQINESA
                     IESLQGVKGQFKVAGQYQIIFGSGIVNQVYAEMAKLTGLVEMSTNDVASAGAEKQNWA
                     QRAVKGLSDIFVRIIPAIVAGGLLMGLFNVLTAPGLFIEGQSLIDANPGLADLASMIN
                     TFANAPFVYLPVLLAFSASRKFGGNPFLGAALGMLMVHPDLLNGWGFGSASVSGTVPT
                     WNILGFEIEKVGYQGSVLPVLVSAYILAKIENGLRKIVPSVVDNLLTPMLAIFITGFL
                     TFTVVGPLTRDVGFMLGDALNWLYDSAGFVGGALFGFIYAPFVITGMHHSFIAIETQL
                     LADIVTTGGTFIFPIAAMSNIAQGAAALAVGVMTKETKLKGVAIPSGVTALLGITEPA
                     MFGVNLKLRYPFIAAICGAALASAFITLFNVKAQALGAAGLPGIISINPQQIGYYIMG
                     MAISFVAAFALTVLLAMREKTKQAAQVTA"
     gene            1854. .2846
                     /gene="VCA0654"
     CDS             1854. .2846
                     /gene="VCA0654"
                     /note="similar to GB:L14685 PID:289864; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sucrose operon repressor ScrR, putative"
                     /protein_id="AAF96555.1"
                     /db_xref="GI:9658070"
                     /translation="MASLHDVARLAGVSKSTVSRVINDEYGVKESTKIKVLKAVEQCG
                     YMVNQVAKDLKSQKTNLIGVIVPRVSSHATAQGVDGLTAIFEQAGKHVLLANTHQVHA
                     KEIEYIQIFNQKRVEGIVFYATHLDQPLVHAIQHSAVPVVLVGQDGSLYNIPSIVHDD
                     HRVGFEAGQRLVKAGCKQIGFIGVQSDDIAVDVLRSQGLQQALGFHQQTLLFHARGNF
                     SIESGYQMAKQAYLDYPQLDGLFCATDRLAVGAIKALQELGVHVGQQVKLLGVGNDEL
                     AYVSNPPLSTFNYAFDKAGENAAKMLLERIAGRGQEMSKVVLTFQNIQRETCPV"
     gene            2922. .4568
                     /gene="VCA0655"
     CDS             2922. .4568
                     /gene="VCA0655"
                     /note="similar to GB:M74035 PID:155269; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sucrose-6-phosphate dehydrogenase"
                     /protein_id="AAF96556.1"
                     /db_xref="GI:9658071"
                     /translation="MLLDTLLELAGGINNVTRILAPQGQVVLALKHPPLVPHLPDDVS
                     LQSVLGEWQLSVQRTAEVSDQQLAAIGKAIAERQKLETLPYQTALDCPYRPLWHISPP
                     QGLLNDPNGFIYHQGEYHLFYQWHPFACEHKDKYWVHLKSLDLVDWQWQSVALTPSDW
                     FDSHGVFSGHAVSHQQDLWLFYTGNTRLGVDRQRQTMQCAARMNANGEFEKLGPVIRC
                     LPEGVTEHIRDPKVIYTQGKWQMLLGAQTLAHQGRLAVYHSDDLLHWHFDKLYGDELG
                     DYGYMWECPDWFELQGEAFFVFGPQGIASANPHHTIEHQNRIFRATQNAQGEIALLQG
                     WPLDEGFDFYAPQTAQTADGRRVLCGWMGLPDETQHPSCDQGWIHQLTALRELEWREG
                     RIYQHPLRELDTLQSEPHTLLLSDNVTELKTKSFALQVTLPWGCELRLMQNTQYRVTL
                     TLDAENQLLRLDRSATQIRQGDTIRELKLDSPTVELRILADQSSLEIFINQGEHVMTS
                     RIFTPRDASGISLHGASVDAKLYYMAPASAPFNLEVNVQP"
     gene            complement(4690. .5661)
                     /gene="VCA0656"
     CDS             complement(4690. .5661)
                     /gene="VCA0656"
                     /note="similar to GB:M76768 SP:P22824 PID:155263;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="fructokinase"
                     /protein_id="AAF96557.1"
                     /db_xref="GI:9658072"
                     /translation="MNALVRLSSNHILRSDSMSRVWLTGDAVVDLIPDGQQHYLKCPG
                     GAPANVAVAIARLCGRSAFFGRVGNDPFGRFMQQTLTDEQVDCQHLYFDPVHRTSTVV
                     VDLDEHGERSFTFMVKPSADQFLQLSDIPSFQNGEWLHVCSIALANQPSRSSTFAAIA
                     QMKEVGGYVSFDPNLREEVWSEPQELQATVMRAVGLADVVKFSEEELQFLTGTQSIEE
                     GLQAIADFQIPLVVVTLGAKGALVVTPNSRQIVSGKAVKPIDTTGAGDAFVGGLLYRL
                     SVAQDWHNQATILDAVKWANGCGALATTQKGAMTALPNQAALYAFLE"
     gene            6008. .7546
                     /gene="VCA0657"
     CDS             6008. .7546
                     /gene="VCA0657"
                     /note="similar to SP:P13035 GB:M21277 GB:M55989 GB:M96795
                     PID:146181; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="aerobic glycerol-3-phosphate dehydrogenase"
                     /protein_id="AAF96558.1"
                     /db_xref="GI:9658073"
                     /translation="MNARSTRFTQDEVLDLIIVGGGINGAGIAADATGRGLKVGLYDA
                     KDFASATSSASSKLVHGGLRYLEHYEFRLVSEALAEREVLLKKAPHIVTPMRFRLPHR
                     PFLRPAWMIRAGLFLYDNLGKRTSLPASHKVNLKAGSVTKPEMQIGFEYSDCWVDDAR
                     LVILNAMQAQEQGAEVLNYCTVEKAERMGDLWHVTLLDEQTQQRFERRSHALVNAAGP
                     WVKQFLNENAHVSSPYGIRLIQGSHIIVPRIHDEPQAYILQNEDKRIVFVIPYLDDYS
                     MIGTTDVEYKGDPRKVAITDAERDYLISIVNKHFMREIARSDIIAEFSGVRPLCDDES
```

```
                NSPQAITRDYTLSLDQQADEAPLLSIFGGKLTTYRKLGEAAMKHLAPFFPKMKAPWTA
                DAPLPGGENFDYAALKNQLVAAFPFITESVIERWLRSYGSRTTQLLAGVTGIEDLGIA
                FSGELYQREIDYLCEKEFARHAQDIFWRRSKLGLNHDTSVVEEVESYLQQKFHAEQPL
                KATV"
     gene       complement(7648. .9258)
                /gene="VCA0658"
     CDS        complement(7648. .9258)
                /gene="VCA0658"
                /note="similar to GB:X61367 PID:48195; identified by
                sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="methyl-accepting chemotaxis protein"
                /protein_id="AAF96559.1"
                /db_xref="GI:9658074"
                /translation="MYTFALIKITLEQQAEMSAYTPSAQQEVLVGDHDQLVSTTDLKG
                VITYCNDTFCRIAGFQADELLGKNHNIVRHASMPKAAFADMWHHLKQGHAWRGIVKNR
                TKSGGFYWVDAYVTPIYQQGQLTGYQSVRVKAERKWVEIATKAYQALLAAEKAGKKIQ
                FKLHTSLRYALLLGALMSPALAHGFQAPEQWQWLASLLPAGVLGLLFRQELVRTPQQL
                KQWQNEYDSISRLIYSGADAFSVADYHLKMASARIRTILGRMMDSARPLGELANQLHL
                TTQEVHQALAAQNSNIQAVTQATDAVESAAERVSSHTHSAHQLIDQVQDHCAETKHSI
                NVTHQNLQRLATQAESAALTTLKLSDQAQQVGQLMTEIGGIAEQTNLLALNAAIEAAR
                AGEQGRGFAVVADEVRALSARTQRATQQIQTSIDTMLSTIEAWRGDITASRDQTEQCA
                QDANTTLQQLQDVECVMSDMLRVIGEVASAAQHQRELTCEVNQHIHSIASVATQNSAA
                THTVEQLAMAMSGKVAEFGALSKQFAQK"
     gene       complement(9444. .10067)
                /gene="VCA0659"
     CDS        complement(9444. .10067)
                /gene="VCA0659"
                /note="similar to PID:2062640; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="protein F-related protein"
                /protein_id="AAF96560.1"
                /db_xref="GI:9658075"
                /translation="MFEMKFWVLCVTFLLTGCGSLVTDRMFNDNLLDVAPKGDFDVRY
                PEWGYAPQKNTSVMPNAMTQPRSASYEDLRQYLLSNGIQHEVVPGDYPMIKLNNTVRF
                ETGSAKVSMASKQWLDTVARFLATESGIDIVLEGHTDNTGSEKLNDKLAERRANAVKA
                ALVQSRVAQNAIYTRGFGENVPACTNSTKNGRACNRRVEIRFILASN"
     gene       complement(10068. .10190)
                /gene="VCA0660"
     CDS        complement(10068. .10190)
                /gene="VCA0660"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF96561.1"
                /db_xref="GI:9658076"
                /translation="MNGEHSNHGLCDGKQPICKLLLMTDKVFRNMPINYKKFNN"
BASE COUNT     2711 a   2569 c   2371 g   2613 t
ORIGIN


  Query Match           15.2%;  Score 231.8;  DB 1;  Length 10264;
  Best Local Similarity  55.4%;  Pred. No. 3.5e-52;
  Matches  497;  Conservative    0;  Mismatches  387;  Indels   13;  Gaps    2;

Qy       3 catggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcg 62
           | | ||   ||| || || |  |  ||||| || ||| | |  ||  | ||  |
Db    1071 CTTCGCGAATGCCCCTTTTGTTTATTTGCCTGTATTGTTGGCATTTTCTGCCTCGCGCAA 1012

Qy      63 tttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagct 122
             || ||||| ||    ||  ||||||| |||  | ||||||  |||||||   ||
Db    1011 GTTTGGCGGTAACCCCTTTTTGGGCGCTGCG-CTGGGTATGCTGATGGTGCATCCCGATC 953

Qy     123 tggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtccc 182
           |  |||||||||  |   | | |    |    |  | |||    | ||||  |||
Db     952 TCTTGAACGGCTGGGGTTTTGGCAGTGCATCCGTATCTGGCACCGTTCCAACTTGGAATA 893

Qy     183 tgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtgg 242
           |  |||||||  ||  | |   |||  ||||| || ||| | ||  | || ||||||||

Db     892 TTCTTGGTTTTGAAATCGAAAAAGTGGGTTATCAAGGCTCGGTATTGCCCGTGCTGGTTT 833

Qy     243 tttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcag 302
             |||  |||||||| |. || || ||   | | | |||      |   |      | |
Db     832 CCGCTTACATTCTGGCCAAAATTGAAAATGGCTTACGCAAAATCGTACCGTCTGTAGTGG 773

Qy     303 acttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcca 362
           |   | || | || ||| ||||| || |  |  |||| || || || || || | |
Db     772 ATAACTTGCTTACCCCAATGCTGGCGATTTTTATCACGGGTTTTCTGACCTTTACTGTGG 713

Qy     363 ttggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttttatg 422
           ||||  ||   || |||    || |||    || ||| | | |  ||  |  | || ||||
Db     712 TTGGACCGCTAACGCGTGATGTCGGCTTCATGTTGGGAGATGCACTGAACTGGCTGTATG 653

Qy     423 atttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatca 482
           | |  |  ||    ||||| ||    || || ||  |  ||||  | ||  | || ||||
Db     652 ACTCGGCAGGGTTTGTCGGTGGCGCACTGTTTGGCTTTATCTATGCTCCTTTTGTGATCA 593

Qy     483 ctggtctgcaccagtccttcccgccaattgagctggagctgttt-----------aacc 530
           | ||  |||||||   |||     | || ||     ||||  ||            | |
Db     592 CCGGAATGCACCATAGCTTTATCGCGATAGAAACTCAGCTACTTGCCGATATCGTGACCA 533

Qy     531 agggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcat 590
           ||||||  |||| || ||| | |  ||  | ||| |||| || || |||||  || ||
Db     532 CGGGTGGCACCTTTATTTTCCCTATCGCTGCGATGTCTAACATTGCGCAGGGCGCCGCGG 473

Qy     591 gtttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcag 650
           |||  ||  |||    |  || ||||       || ||||| ||  | ||   | |||
Db     472 CTTTAGCTGTGGGTGTGATGACGAAAGAGACCAAACTCAAAGGGGTCGCGATTCCTTCGG 413

Qy     651 gtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgc 710
           ||||  ||||  | || || || || || ||||| || || || |||||  |     |||
Db     412 GTGTGACCGCGCTGCTGGGGATCACTGAACCTGCCATGTTTGGCGTGAATTTGAAATTGC 353

Qy     711 gctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactct 770
           | |  |||||  |    |  ||| ||   || ||| | |   | |||||  ||  |  | |
Db     352 GTTATCCGTTTATTGCTGCCATCTGTGGTGCGGCTTTAGCCAGTGCTTTCATCACTTTAT 293

Qy     771 ttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatg 830
           | ||| | || ||    ||| | ||||| || |||||    ||  |  |||| ||| ||
Db     292 TCAATGTAAAAGCCCAAGCGCTAGGCGCGGCTGGTTTGCCAGGGATCATTTCGATTAATC 233

Qy     831 ctccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcg 887
           | |   | || |     | | |  || ||   |  |  | ||  |   ||| || |
Db     232 CACAGCAAATTGGTTACTACATCATGGGTATGGCGATCTCATTTGTGGCCGCCTTTG 176


RESULT  10
STRSCRA
LOCUS       STRSCRA                 2508 bp    DNA     linear   BCT 21-MAR-2000
DEFINITION  Streptococcus mutans sucrose-6-phosphate hydrolase (scrB) gene,
            partial cds; and enzyme scr-II gene, complete cds.
ACCESSION   M22711
VERSION     M22711.1  GI:153799
KEYWORDS    enzyme scr-II; phosphoenolpyruvate-dependent sucrose
            phosphotransferase system; sucrose-6-phosphate hydrolase.
SOURCE      Streptococcus mutans.
  ORGANISM  Streptococcus mutans
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 2508)
  AUTHORS   Sato,Y., Poy,F., Jacobson,G.R. and Kuramitsu,H.K.
  TITLE     Characterization and sequence analysis of the scrA gene encoding
            enzyme IIscr of the Streptococcus mutans phosphoenolpyruvate-
            dependent sucrose phosphotransferase system
  JOURNAL   J. Bacteriol. 171, 263-271 (1989)
  MEDLINE   89123027
COMMENT     Draft entry and computer-readable sequence for [1] kindly provided
            by H.K.Karamitsu, 22-FEB-1989.
FEATURES             Location/Qualifiers
     source          1. .2508
                     /organism="Streptococcus mutans"
                     /strain="GS-5"
```

```
                     /db_xref="taxon:1309"
                     /clone="pMH613"
     CDS             complement(<1. .28)
                     /note="sucrose-6-phosphate hydrolase (scrB)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26970.2"
                     /db_xref="GI:7272364"
                     /translation="MNLPQNIRY"
     CDS             223. .2217
                     /note="enzyme scr-II"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26971.1"
                     /db_xref="GI:153801"
                     /translation="MDYSKVASEVITAVGKDNLVAAAHCATRLRLVLKDDSKVDQKAL
                     DKNADVKGTFKTDGQYQVIIGPGDVNFVYDEIIKQTGLTEVSTDDLKKIAASGKKFNP
                     IMALIKLLSDIFVPIIPALVAGGLLMALNNFLTSEGLFGTKSLVQQFPIIKGSSDMIQ
                     LMSAAPFWFLPILVGISAAKRFGANQFLGASIGMIMVAPGAANIIGLAANAPISKAAT
                     IGAYTGFWNIFGLHVTQASYTYQVIPVLVAVWLLSILEKFFHKRLPSAVDFTFTPLLS
                     VIITGFLTFIVIGPVMKEVSDWLTNGIVWLYDTTGFLGMGVFGALYSPVVMTGLHQSF
                     PAIETQLISAFQNGTGHGDFIFVTASMANVAQGAATFAIYFLTKDKKMKGLSSSSGVS
                     ALLGITEPALFGVNLKYRFPFFCALIGSASAAAIAGLLQVVAVSLGSAGFLGFLSIKA
                     SSIPFYVVCELISFAIAFAVTYGYGKTKAVDVFAAEAAVEEAIEEVQEIPEEAASAAN
                     KAQVTDEVLAAPLAGEAVELTSVNDPVFSSEAMGKGIAIKPSGNTVYAPVDGTVQIAF
                     DTGHAYGIKSDNGAEILIHIGIDTVSMEGKGFEQKVQADQKIKKGDVLGTFDSDKIAE
                     AGLDNTTMFIVTNTADYASVETLASSGTVAVGDSLLEVKK"
     misc_signal     2227. .2252
                     /note="pot. transcription termination signal; putative"
BASE COUNT      740 a    401 c    513 g    854 t
ORIGIN


  Query Match           15.1%;  Score 230.8;  DB 1;  Length 2508;
  Best Local Similarity   50.1%;  Pred. No. 5.3e-52;
  Matches  779;  Conservative    0;  Mismatches  717;  Indels   58;  Gaps    6;

Qy      2 tcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagc 61
          | ||| || | ||||| |||  |||||||||  | ||||||||  |  | ||  | || |
Db    704 TAATGTCAGCAGCGCCTTTCTGGTTCTTGCCCATTTTGGTTGGGATTTCAGCCGCTAAAC 763

Qy     62 gtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagc 121
          |||| || |  ||  | ||| | || ||  |  || || |||  |||||    ||| |
Db    764 GTTTTGGTGCTAACCAATTCTTAGGAGCTTCA-ATCGGCATGATTATGGTCGCACCGGGG 822

Qy    122 ttggtgaacggctacgacgtggccgcca----------------------ccatggctgcg 160
            |  ||   |   |  | || || |                        | |    ||
Db    823 GCAGCAAATATCATTGGTTTAGCAGCTAATGCTCCAATTTCAAAGGCTGCTACTATTGGT 882

Qy    161 ggcgaaatgccaatgtggtccctgtttggtttagatgttgcccaagccggttaccagggc 220
          |   | |    | | |||    | ||||| ||  ||||| | || ||  ||||
Db    883 GCTTATACAGGATTTTGGAATATTTTTGGGTTGCATGTTACTCAGGCAAGTTATACTTAT 942

Qy    221 accgtgcttcctgtgctggtggtttcttggattctggcaacgatcgagaagttcctgcac 280
             ||| |||| || || |||| |   ||| || |  | |   | || || ||  | ||
Db    943 CAAGTGATTCCAGTACTTGTGGCTGTATGGCTTTTATCTATCTTAGAAAAATTTTTCCAT 1002

Qy    281 aagcgactcaagggcactgcagacttcctgatcactccagtgctgacgttgctgctcacc 340
          ||| || |        | |  |||||     | || ||  |  || |  |  |  | ||
Db   1003 AAGAGATTGCCCTCAGCAGTTGACTTTACCTTTACACCTTTATTGTCAGTTATTATTACT 1062

Qy    341 ggattccttacattcatcgccattggcccagcaatgcgctgggtgggcgatgtgctggca 400
          |||||  | ||||| ||||  ||||| || |  |||       ||    ||  |||  ||
Db   1063 GGATTTTTGACATTTATCGTTATTGGTCCTGTTATGAAAGAAGTTTCTGACTGGCTTACA 1122

Qy    401 cacggtctacagggactttatgatttcggtggtccagtcggcggtctgctctcttcggtctg 460
           | ||  |      | || ||||||       ||      | ||      ||||  |||
Db   1123 AATGGAATCGTATGGCTGTATGATACAACGGGATTCCTAGGTATGGGTGTCTTTGGTGCT 1182

Qy    461 gtctactcaccaatcgtcatcactggtctgcaccagtccttcccgccaattgagctggag 520
          | || |||||  | || || |||||||| || ||    |||||  | |||||      |
Db   1183 CTGTATTCACCTGTAGTTATGACTGGTCTCCATCAAAGTTTCCCAGCTATTGAAACCCAA 1242

Qy    521 ctgtt------------------taaccagggtggatccttcatcttcgcaacggcatct 562
          ||  |                 ||       |||  ||||| |  || || ||
Db   1243 CTCATTTCAGCTTTCCAAAATGGTACTGGTCATGGTGACTTTATCTTTGTTACAGCTTCA 1302

Qy    563 atggctaatatcgcccagggtgcggcatgtttggcagtgttcttcctggcgaagagtgaa 622
          ||||| ||| |||| || || || |||    || ||| | | |||| |  | ||   | |
Db   1303 ATGGCCAATGTCGCACAAGGCGCAGCAACCTTTGCAATTTACTTCTTAACAAAAGATAAG 1362

Qy    623 aagctcaagggccttgcaggtgcttcaggtgtctccgctgttcttggtattacggagcct 682
          ||| | || || ||  ||  | ||||||||| || ||  |||||||||||| || |||
Db   1363 AAGATGAAAGGTCTCTCATCTTCTTCAGGTGTGTCAGCGCTTCTTGGTATTACAGAACCT 1422

Qy    683 gcgatcttcggtgtgaaccttcgcctgcgctggccgttcttcatcggtatcggtaccgca 742
          ||  | || || |||||  |        || |  || |||||    | | |   |  | ||
Db   1423 GCTCTTTTTGGGGTGAATTTGAAATATCGTTTCCCATTCTTTTGTGCTCTGATTGGCTCA 1482

Qy    743 gctatcggtggcgctttgattgcactctttaatatcaaggcagttgcgttgggcgctgca 802
          ||||  | ||  ||  |    |   | || | |    || ||| |  | ||  | |||
Db   1483 GCTAGTGCTGCAGCAATTGCGGGCTTACTTCAAGTTGTAGCGGTTTCTCTTGGTTCGGCA 1542

Qy    803 ggtttcttgggtgttgtttctattgatgctccagatatggtcatgttcttggtgtgtgca 862
          |||||  | ||| |  |||||||| | ||       |||      | |   | || |||| |
Db   1543 GGTTTTCTTGGTTTCCTTTCTATTAAGGCAAGTTCTATTCCATTTTATGTAGTCTGTGAA 1602

Qy    863 gttgttaccttcttcatcgcattcggcgcagcgattgcttatggcctttacttggttcgc 922
          |  | |  ||    || || || |  | | |   ||  |||||          || |
Db   1603 TTAATCAGTTTTGCTATAGCTTTTGCAGTAACTTATGGCTATGGTAAGACGAAGGCTGTT 1662

Qy    923 cgcaacggcagcattgatccagatgcaaccgctgctccagtgcctgcaggaacgaccaaa 982
                |      |||  | | ||      |  |||   |      |     |   |    |
Db   1663 GATGTCTTTGCTGCTGAAGCTGCTGTCGAAGAAGCTATTGAAGAAGTGCAAGAAATTCCA 1722

Qy    983 gccgaagcagaagcacccgcagaattttcaaacgattcca---ccatcatccaggcacct 1039
          |  ||||| |   |  | ||| |      || | | | |         |  |  ||| |||
Db   1723 GAAGAAGCTGCTTCTGCAGCAAATAAGGCACAAGTTACTGATGAAGTTCTTGCGGCTCCT 1782

Qy   1040 ttgaccggtgaagctattgcactgagcagcgtcagcgatgccatgtttgccagcggaaag 1099
          |  | |||||||||| ||| | | | |   || |  ||  |  | ||| | |||| |
Db   1783 CTTGCTGGTGAAGCTGTTGAATTAACCTCTGTTAATGACCCTGTTTTTTCTAGCGAAGCA 1842

Qy   1100 cttggctcgggcgttgccatcgtcccaaccaaggggcagttagtttctccggtgagtgga 1159
          | ||   |||  |||| |||   || |              |   | | |||||   |||
Db   1843 ATGGGTAAGGGGATTGCTATCAAACCTAGTGGTAATACAGTTTATGCACCGGTTGATGGG 1902

Qy   1160 aagattgtggtggcattcccatctggccatgctttcgcagttcgcaccaaggctgaggat 1219
          |   ||    | || ||     ||||||||||  |   ||    |  |  ||
Db   1903 ACTGTTCAAATTGCTTTTGATACTGGCCATGCTTATGGTATTAAGTCAGACAATG----- 1957

Qy   1220 ggttccaatgtggatatcttgatgcacattggtttcgacacagtaaacctcaacggcacg 1279
               || ||| ||  | || || |||||| | || ||||||    |  | || |
Db   1958 -------GTGCGGAGATTCTTATTCATATTGGTATTGATACAGTATCAATGGAAGGTAAA 2010

Qy   1280 cactttaacccgctgaagaagcagggcgatgaagtcaaagcaggggagctgctgtgtgaa 1339
             ||| | |          |    |  | || | |||  ||| ||  |  ||  |   |
Db   2011 GGATTTGAACAAAAAGTTCAAGCAGATCAAAAAATTAAAAAAGGTGACGTTCTCGGAACA 2070

Qy   1340 ttcgatattgatgccattaaggctgcaggttatgaggtaaccacgccgattgttgtttcg 1399
          || |||| |||    |||   | || |||  |||     || ||   | |  |||| |
Db   2071 TTTGATAGTGACAAAATTGCAGAAGCTGGTCTTGATAATACAACAATGTTCATTGTTACT 2130

Qy   1400 aattacaagaaaaccggacctgtaaacacttacggtttgggcgaaattgaagcgggagcc 1459
          |||       |    |  | ||  | ||   || ||      | || || || | ||
Db   2131 AATACAGCTGATTATGCGTCAGTTGAAACCCTCGCTTCTTCAGGAACTGTTGCTGTAGGT 2190

Qy   1460 ---aacctgctcaacgtcgcaaagaaagaagcggtgccagcaacaccataagtt 1510
             | | |  ||  | ||   |||  |||||     |  ||| || ||| |||| |
Db   2191 GATAGCTTACTTGAAGTTAAAAAATAAGAAATATTATCAGAAAGACCGTAAGGT 2244


RESULT  11
AE006222/c
LOCUS       AE006222               10085 bp    DNA     linear   BCT 08-MAR-2001
DEFINITION  Pasteurella multocida PM70 section 189 of 204 of the complete
```

```
            genome.
ACCESSION   AE006222 AE004439
VERSION     AE006222.1  GI:12722266
KEYWORDS    .
SOURCE      Pasteurella multocida.
  ORGANISM  Pasteurella multocida
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Pasteurella.
REFERENCE   1  (bases 1 to 10085)
  AUTHORS   May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
            Kapur,V.
  TITLE     Complete genomic sequence of Pasteurella multocida,Pm70
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
  MEDLINE   21145866
REFERENCE   2  (bases 1 to 10085)
  AUTHORS   Zhang,Q. and Kapur,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
            University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
            55108, USA
FEATURES             Location/Qualifiers
     source          1. .10085
                     /organism="Pasteurella multocida"
                     /strain="PM70"
                     /db_xref="taxon:747"
     gene            344. .625
                     /gene="PM1841"
     CDS             344. .625
                     /gene="PM1841"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03925.1"
                     /db_xref="GI:12722267"
                     /translation="MVNCRIGSLEILHKQLGHQSIVNCRIGSLEKDSRAGRKSYSVNC
                     RIGSLESLCIAFFNSTYVNCRIGSLEKNQNLHLFHLDVNCRIGSLERIY"
     gene            1158. .1400
                     /gene="PM1842"
     CDS             1158. .1400
                     /gene="PM1842"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03926.1"
                     /db_xref="GI:12722268"
                     /translation="MQCASFNTDRVNCRIGSLEIFSQDCLLNPDVNCRIGSLENQKNC
                     IAKCFIVNCRIGSLEMNNVFRRNATFVNCRIGSLEM"
     gene            complement(2259. .3224)
                     /gene="PM1843"
     CDS             complement(2259. .3224)
                     /gene="PM1843"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03927.1"
                     /db_xref="GI:12722269"
                     /translation="MKICLVKTSSMGDILHSLPALTDAQRALPNLQVDWVVEESFAEI
                     PHWHSAVRQVIPIALRRWRKTPFTALTRKQWKTYRTLLQSEQYDAVIDAQGLLKSALF
                     ATRLARGTRHGYDRHSIREPIAACFYDKKYAISYQQHAVERIRQLFAQSLNYPLPHTQ
                     GDYGIARHFPQHAQTMPYVLFFHGTTRQDKFLPDRQWRTLAEKLTAQGIQIHVPWSNS
                     QEKQRAERIAQGLDLVTILPKSTLSELAQQIVNASAVVSVDTGLAHLSAALDKTNLTL
                     YGATDPKLIGSYGKNQHYLVAESMDKIEADEIFQKLTALLSLNTI"
     gene            complement(3221. .4267)
                     /gene="rfaF"
                     /note="PM1844"
     CDS             complement(3221. .4267)
                     /gene="rfaF"
                     /codon_start=1
                     /transl_table=11
                     /product="RfaF"
                     /protein_id="AAK03928.1"
                     /db_xref="GI:12722270"
                     /translation="MHILIIGPSWVGDMMMSHSLYQTLKQHYPLCEIDVMAPNWCKPL
                     LERMPEVRQAITMPLGHGAFALGDRYRLGKALRHQYDMAIVLPNSLKSAFIPFFAKIP
                     LRRGWKGESRYLFLNDLRSNKKDYPMMVQRYIALAFEKGAVPKAEDLPRPVPYLTVDP
                     VTQQHTLKKFEKQTALISPRPSIGFCPGAEFGPAKRWPHYHYAKLAEMLIQQGYAIRL
                     FGSAKDEAVGEEIRQALPETMRHFCINLAGQTSLNEAVDLIADCTAIVTNDSGLMHIA
                     AAVQRPLVALYGPTSPTYTPPLSEKAVIIRLIEGGLIKVRKGKDSDEGYHQSLIDIQP
                     ELVFDKLTGLLANT"
     gene            complement(4333. .4743)
                     /gene="PM1845"
     CDS             complement(4333. .4743)
                     /gene="PM1845"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03929.1"
                     /db_xref="GI:12722271"
                     /translation="MKSLTQLLFILTALLPTSLFAHIHVEQAQMFSAKAGEPSAIFMN
                     IHNTGNEDVSLALVQSDIRADIVLHGTQHGKMIEVTGINLPAHQMTALKRGGLHIMVF
                     NVEQDLNVGDSFPLRLLFDNGEIMHIEAKIIQYQ"
     gene            complement(4859. .6283)
                     /gene="ptsB"
                     /note="PM1846"
     CDS             complement(4859. .6283)
                     /gene="ptsB"
                     /codon_start=1
                     /transl_table=11
                     /product="PtsB"
                     /protein_id="AAK03930.1"
                     /db_xref="GI:12722272"
                     /translation="MNFSQIAQQVIDKLGGKDNISAAAHCATRLRIVVKNENLIDKKG
                     IENIEGVKGQFAVAGQYQIIFGSGTVNKVYAALSKLLGIGDMTTSEVAAAGTEKQGLL
                     QRLVKGLADIFVPIIPAIVAGGLLMGIHSMLTAKGFFVEEKNVVDLYPAIADLVDFIN
                     TIANAPFVFLPVLLGFSATRKFGGNPFLGAALGMLLVHPALSDGWNYALTLAKGNIQY
                     WHIFGLEIERVGYQGTVIPVLVASWVLATLEKNLRKVVPSFLDNLITPLFALFITGLL
                     AFTVIGPIGREAGSLISTGLTWLYDTLGFVGGAIFGTLYAPIVITGMHQTFIAVETQL
                     LAEVARTGGTFIFPIAAMSNIAQGAACLGAAYVMKDAKVRGIAVPSGISALLGITEPA
                     MFGVNLRYRYPFISAMIGAGISSAVIALFNVKAIALGAAGLPGIPSIKPDSLAMYCVG
                     MLISASIAFTLTVILGKRAQLKAE"
     gene            complement(6512. .7531)
                     /gene="scrR"
                     /note="PM1847"
     CDS             complement(6512. .7531)
                     /gene="scrR"
                     /codon_start=1
                     /transl_table=11
                     /product="ScrR"
                     /protein_id="AAK03931.1"
                     /db_xref="GI:12722273"
                     /translation="MSITNKKRITLSDIAKCCGLSTTTVSMILNNRGEEFKIKAETCQ
                     RVKAIAEQQGYRANVYAKSLKMQRSNVIGFVIPDLTNYGFACTAKILEKQCRENGLQL
                     IIACSDDNPQQEKLAIERLLDRQIDFLITAPTHQDPSYYQKIGQHTPILQLDRYVPDL
                     ALNYVISNDSEHITALICRLCDQYQLTEFMYFGGQRTLSPSESRLAGFLQGLKQANLN
                     LKKEWILHRTYQPESGYQMLAETEAKLGRLPQAIFTASYTLLEGVLRYLTEQKKMDLL
                     LNRQLHLATFDDHDILNALPFHIHSIKQNHPQLAHQIFQHIMRYFAGKKSPNQKIDCD
                     VIYRD"
     gene            complement(7557. .9056)
                     /gene="scrB"
                     /note="PM1848"
     CDS             complement(7557. .9056)
                     /gene="scrB"
                     /codon_start=1
                     /transl_table=11
                     /product="ScrB"
                     /protein_id="AAK03932.1"
                     /db_xref="GI:12722274"
                     /translation="MPFSLSKVRSFFCHFWTSPMYIFNQGKYKSLHAQQVGELAQIRQ
                     HVLRDTQFRPTYHLAPETGLLNDPNGLVFDGEKYHIFYQWFPYAAIHGMKHWKHFMTY
                     DFHHFQVADPLIPDELFESHGCYSGGALLWQDQIVAFYTGNTRNAENQRIPHQNIAIF
                     KKDGTLLKKRCLLNQAPTGYTEHVRDPKPFITEEGKIRFILGAQREDLSGTALIYEMT
                     DLDATPQLIGEFKLPDFDNQQVFMWECPDLLKLQEQDLFLWSPQGKAQKTYCFQNHYH
                     ATYALGQYANLHFSAQHIEELDYGFDFYAPQSIQGSEGIYYAWVGLPDLTYPTDEYQW
                     HSMLSLPRQFSIQQGILYQRPIPQVYAQCLPAHKYWVEGKQQIPHLDRSYVAFDAENQ
                     PFTLRFFTNQRGQQLILRYQNGILTLDRTYTTQTALMQQFGEQRHCILPQLNHVEIFF
                     DRSVVEIFLNQGEKVMTSRFFIADRENALDSSRPLCVSVAELKPIILTP"
```

```
     gene            complement(9037. .9963)
                     /gene="PM1849"
     CDS             complement(9037. .9963)
                     /gene="PM1849"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAK03933.1"
                     /db_xref="GI:12722275"
                     /translation="MGKIWVLGDAVVDLIPEGENTYLRCAGGAPANVAVGIARLGATT
                     GFIGRVGDDPLGRFMLNKLQQENVCTQYMHLDPKQRTSTVIVDLKEGERSFTFMVNPS
                     ADQFLELSDLPHFNPNDWLHCCSIALINEPSRSSTFAAIHRIKAEGGFFSFDPNLRES
                     LWNSPAEMKTVVNQVVALADVLKFSEEELLFLTDSDTLSDATAKLTQHYPDKLIIVTL
                     GKLGATYYLQGKSQLVSSKKALIPVDTTGAGDAFVSGLLFGISQVKNWQNPTALVEII
                     KQANACGALATTAKGAMSALPNKQQLDAFLAE"
BASE COUNT     2923 a   1988 c   2212 g   2962 t
ORIGIN


  Query Match             14.3%;  Score 218.6;  DB 1;  Length 10085;
  Best Local Similarity   54.5%;  Pred. No. 1.5e-48;
  Matches  488;  Conservative    0;  Mismatches  394;  Indels   13;  Gaps    2;

Qy      3 catggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcg 62
          ||| |||  ||| || || |  ||||| || || ||  |||||||  |||| ||
Db   5801 CATTGCAAATGCACCTTTTGTCTTCTTACCTGTATTACTTGGTTTTTCCGCCACTCGAAA 5742

Qy     63 tttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagct 122
           || ||||| |||   ||| | || || ||   |||| |||    | |||  |||
Db   5741 ATTTGGCGGTAATCCATTCTTAGGAGCGGCAC-TTGGCATGCTATTAGTGCACCCAGCAC 5683

Qy    123 tggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtccc 182
           |     | ||||   |    ||    ||  | |||   ||| | || | |   |||
Db   5682 TTTCTGATGGCTGGAATTATGCACTGACACTCGCTAAAGGCAATATTCAATATTGGCATA 5623

Qy    183 tgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtgg 242
           | ||||||||||  |||  |  | ||| ||||| || || ||  | || || || || |
Db   5622 TCTTTGGTTTAGAAATTGAACGTGTCGGCTACCAAGGTACAGTTATCCCCGTTCTTGTCG 5563

Qy    243 tttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcag 302
            |||||| |  | |||||  | |||||   ||| |  ||      |   |     |  |
Db   5562 CCTCTTGGGTCTTAGCAACCTTGGAGAAAAACCTACGTAAAGTCGTTCCGTCATTTCTTG 5503

Qy    303 acttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcca 362
           |   | | || || ||| |  |  | ||  |  | || || || ||  | || | ||  |
Db   5502 ATAACTTAATTACCCCACTTTTGCATTATTTATAACGGGCTTACTCGCTTTTACCGTAA 5443

Qy    363 ttggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatg 422
           | || ||      |||    | |||   | |  |       ||  |     |  | ||||
Db   5442 TAGGTCCTATCGGGCGTGAAGCGGGGTCTTTAATTAGTACGGGCTTGACCTGGTTGTATG 5383

Qy    423 atttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatca 482
           |       ||    || || |||    | ||||||    | ||| |||| || || ||||
Db   5382 ACACATTAGGATTCGTGGGGGGTGCCATTTTCGGTACACTTTACGCACCCATTGTTATCA 5323

Qy    483 ctggtctgcaccagtcctt-----------cccgccaattgagctggagctgtttaacc 530
           | ||| |||||||  | ||             | | | |       ||| ||  |  |
Db   5322 CAGGTATGCACCAAACATTTATCGCTGTCGAAACACAATTGCTTGCTGAGGTGGCTCGCA 5263

Qy    531 agggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcat 590
            || ||  |||| ||||| |||  ||  | ||| | ||||| || || |||||  || |
Db   5262 CTGGAGGTACCTTTATCTTCCCAATCGCCGCGATGTCAAATATTGCTCAAGGTGCAGCTT 5203

Qy    591 gtttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcag 650
           ||||||  |     |   ||  |||   ||  ||  | |  ||  | || |   |||| |
Db   5202 GTTTGGGCGCCGCTTATGTGATGAAAGATGCCAAAGTAAGAGGGATCGCCGTACCTTCTG 5143

Qy    651 gtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgc 710
           |  |||| ||  |  | || ||||| ||||| || || |||||||| ||| | |||   |
Db   5142 GCATCTCAGCCTTGTTAGGGATTACTGAGCCCGCCATGTTCGGTGTCAACTTACGCTATC 5083

Qy    711 gctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactct 770
           |||  || ||| |    |  ||       || |  ||     | ||  | |||||  | |
```

```
Db   5082 GCTATCCTTTCATATCTGCCATGATCGGTGCTGGCATTTCCAGTGCGGTCATTGCCTTGT 5023

Qy     771 ttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatg 830
           ||||| | || ||  |||| || |||||||| || ||   |||| |    || ||| |
Db    5022 TTAATGTTAAAGCCATTGCTTTAGGCGCTGCGGGCTTGCCGGGTATCCCCTCAATTAAGC 4963

Qy     831 ctccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcatt 885
           |     | | |  ||||  |  ||  ||    || || ||   |  || |||||
Db    4962 CGGATAGTTTAGCAATGTATTGTGTCGGTATGCTTATTTCCGCATCGATTGCATT 4908


RESULT  12
VIBSCRAK
LOCUS       VIBSCRAK                2955 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  Vibrio alginolyticus enzyme II-sucrose protein (scrA) gene and
            fructokinase (scrK) gene, complete cds.
ACCESSION   M76768 M30194
VERSION     M76768.1  GI:155261
KEYWORDS    enzyme II-sucrose protein; fructokinase; sucrase-encoding; sucrose
            uptake-encoding.
SOURCE      Vibrio alginolyticus DNA.
  ORGANISM  Vibrio alginolyticus
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 2955)
  AUTHORS   Blatch,G.L., Scholle,R.R. and Woods,D.R.
  TITLE     Nucleotide sequence and analysis of the Vibrio alginolyticus
            sucrose uptake-encoding region
  JOURNAL   Gene 95, 17-23 (1990)
  MEDLINE   91071601
FEATURES             Location/Qualifiers
     source          1. .2955
                     /organism="Vibrio alginolyticus"
                     /db_xref="taxon:663"
     gene            393. .1832
                     /gene="scrA"
     CDS             393. .1832
                     /gene="scrA"
                     /codon_start=1
                     /transl_table=11
                     /product="enzyme II-sucrose protein"
                     /protein_id="AAA27555.1"
                     /db_xref="GI:155262"
                     /translation="MNYPAVAKELLTLLGGKSNITALAHCATRLRLAVADEQKIDEQA
                     IDNLEGVKGQFKVAGQYQIIFGSGIVNQVYAEMAKLTGMSEMSTNDVASAGAEKQNIV
                     QPAVKGLSDIFVPIIPAIVAGGLLMGIYNLLTAQGLFIDGKSLIEANPGLTDLANMIN
                     TFANAPFVYLPILLAFSASKKFGGNPYLGAALGMLMVHPDLLNGWGFGGASVSGNIPV
                     WNILGFEIQKVGYQGSVLPVLVSAFILAKVELGLRKVIPSVLDNLLTPLLAIFIAGLL
                     TFTVVGPFTRDIGFLLGDGLNWLYNTAGFVGGAVFGLIYAPFVITGMHHSFIAIETQL
                     LADIATTGGTFIFPIAAMSNVSQGAAALAVGVMSKDKKMKGIAIPSGVTGLLGITEPA
                     MFGVNLKLRYPFIAAVCAAALSSAFITMFNVKAQALGAAGLPGIISITPDKIGYYIAG
                     MVIAFLTAFVLTIVLGIGDRAKVGKKAAA"
     gene            1998. .2921
                     /gene="scrK"
     CDS             1998. .2921
                     /gene="scrK"
                     /codon_start=1
                     /transl_table=11
                     /product="fructokinase"
                     /protein_id="AAA27556.1"
                     /db_xref="GI:155263"
                     /translation="MNQVWVTGDAVVDLIPESETSLLKCPGGAPANVAVAIARLSGKS
                     AFFGRVGDDPFGRFMQSILDQEGVCTEFLIKDPEQRTSTVVVDLDDQGERSFTFMVKP
                     SADQFMSVEDMGNFKQGDWLHVCSISLANEPSRSSTFEAIKRAKAAGGFISFDPNLRD
                     EVWQDQSEIQAVVMKAVAMADVVKFSEEELLFLTDSTSMAQGLQQIAAMNIALVLVTQ
                     GAKGVWRVFESQSELITGQVVSPIDTTGAGDAFVGGLLACLSRHADWKNHPVVSSAIQ
                     WANGCGALATTQKGAMTALPTQTELLRFIGQ"
BASE COUNT      723 a    605 c    747 g    880 t
ORIGIN


  Query Match             14.1%;  Score 215.2;  DB 1;  Length 2955;
  Best Local Similarity   53.9%;  Pred. No. 1e-47;
  Matches  493;  Conservative    0;  Mismatches  408;  Indels   13;  Gaps    2;
```

```
Qy       5 tggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcgtt 64
Db     877 TTGCAAACGCGCCGTTTGTTTACCTGCCTATTCTGCTGGCATTTTCAGCCAGCAAGAAGT 936

Qy      65 tcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagcttg 124
Db     937 TTGGCGGCAACCCATATTTAGGGGCAGCG-CTAGGTATGTTGATGGTTCACCCTGACTTG 995

Qy     125 gtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtccctg 184
Db     996 CTGAATGGTTGGGGATTCGGCGGCGCATCTGTGTCTGGCAACATTCCGGTTTGGAATATT 1055

Qy     185 tttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtggtt 244
Db    1056 CTTGGTTTTGAAATCCAAAAAGTGGGTTACCAAGGTTCAGTGTTACCGGTTTTGGTATCG 1115

Qy     245 tcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgcagac 304
Db    1116 GCATTCATTCTTGCGAAGGTTGAGCTAGGCTTACGTAAAGTCATACCTTCTGTTTTAGAC 1175

Qy     305 ttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgccatt 364
Db    1176 AACCTTTTAACGCCATTACTGGCGATTTTCATAGCGGGTTTGTTGACGTTTACTGTCGTC 1235

Qy     365 ggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttatgat 424
Db    1236 GGTCCGTTCACGCGTGATATCGGTTTCCTATTGGGTGATGGTCTGAACTGGCTGTACAAC 1295

Qy     425 ttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatcact 484
Db    1296 ACAGCTGGCTTTGTTGGTGGTGCGGTATTTGGTCTTATTTATGCTCCGTTTGTTATTACT 1355

Qy     485 ggtctgcaccagtccttcccgccaattgagctggagctgttt----------aaccag 532
Db    1356 GGTATGCATCACAGCTTCATTGCGATTGAAACTCAACTTCTTGCGGATATTGCAACGACG 1415

Qy     533 ggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgt 592
Db    1416 GGCGGCACCTTTATCTTTCCTATTGCGGCGATGTCTAACGTTTCTCAAGGAGCGGCAGCG 1475

Qy     593 ttggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggt 652
Db    1476 TTGGCGGTGGGCGTGATGAGCAAAGATAAGAAGATGAAAGGTATTGCGATTCCATCTGGT 1535

Qy     653 gtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgc 712
Db    1536 GTTACGGGTTTGCTGGGTATTACAGAACCAGCCATGTTTGGTGTTAACTTAAAACTACGT 1595

Qy     713 tggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactcttt 772
Db    1596 TACCCATTTATTGCGGCAGTTTGCGCTGCGGCGTTATCAAGTGCGTTCATCACAATGTTT 1655

Qy     773 aatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgct 832
Db    1656 AATGTGAAAGCTCAGGCATTGGGTGCAGCTGGTCTTCCAGGCATCATTTCCATCACGCCA 1715

Qy     833 ccagatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgca 892
Db    1716 GATAAGATTGGGTACTACATCGCGGGTATGGTGATCGCGTTCTTGACCGCATTTGTTCTC 1775

Qy     893 gcgattgcttatgg 906
Db    1776 ACCATTGTTCTTGG 1789


RESULT  13
SXSCRA
LOCUS       SXSCRA                  2655 bp    DNA     linear   BCT 10-FEB-1999
DEFINITION  S.xylosus scrA gene and unidentified open reading frames.
ACCESSION   X69800
VERSION     X69800.1  GI:407905
KEYWORDS    membrane protein; scrA gene; sucrose transport protein.
SOURCE      Staphylococcus xylosus.
  ORGANISM  Staphylococcus xylosus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 2655)
  AUTHORS   Wagner,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1992) E. Wagner, Mikrobielle Genetik,
            Universitaet Tuebingen, Auf der Morgenstelle 28, 7400 Tuebingen 1,
            FRG
REFERENCE   2  (bases 1 to 2655)
  AUTHORS   Wagner,E., Gotz,F. and Bruckner,R.
  TITLE     Cloning and characterization of the scrA gene encoding the
            sucrose-specific Enzyme II of the phosphotransferase system from
            Staphylococcus xylosus
  JOURNAL   Mol. Gen. Genet. 241 (1-2), 33-41 (1993)
  MEDLINE   94049686
FEATURES             Location/Qualifiers
     source          1. .2655
                     /organism="Staphylococcus xylosus"
                     /strain="DSM 20267"
                     /isolate="cured strain C2a"
                     /db_xref="taxon:1288"
                     /clone="pSE4"
                     /clone_lib="sub-genomic in Staphylococcus xylosus mutant
                     scr1"
     CDS             complement(<1. .121)
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA49459.1"
                     /db_xref="GI:407906"
                     /db_xref="SPTREMBL:Q56203"
                     /translation="MKSTTLCIHEFLNTHTQRCIHQVILLFSLNNKLDVTLNGY"
     RBS             complement(132. .137)
     RBS             338. .342
     CDS             350. .925
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA49460.1"
                     /db_xref="GI:407907"
                     /db_xref="SPTREMBL:Q56204"
                     /translation="MQHIYSEFYNQESKITKLKSKVLGIVLHDDVENNSAEDYIVWLQ
                     QRIDNNELEKGWACAYVDKQTCYWFHPSPYVEWHCGNAFANDHYIGIERCQSKINGIL
                     TDEQFGENEEASYWIAALLLKKYQLPVNRETVKLHKMFFDTECPARAWSIHLDNAPTN
                     EENIKMLQDYFIDKIKYHYFNIKESDMFVVG"
     -35_signal      972. .977
     -10_signal      995. .1000
     RBS             1040. .1045
     gene            1053. .2495
                     /gene="scrA"
     CDS             1053. .2495
                     /gene="scrA"
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="EIIscr"
                     /protein_id="CAA49461.1"
                     /db_xref="GI:407908"
                     /db_xref="SWISS-PROT:P51184"
                     /translation="MNYKKSAENILQALGGEDNVEAMTHCATRLRLVLKDEGLVDEKA
                     LGDMDVVKGTFSTGGQYQVIIGSGTVNKVFSELEKITGKEASSVSEVKTQGTKNMNPF
                     QRFVKMLSDIFVPIIPAIVAGGLLMGINNILTAPGIFYDNQSLIEVQNQFSGLAEMIN
                     IFANAPFTLLPILIGFSAAKRFGGNAYLGAALGMILVHPELMSAYDYPKALEAGKEIP
                     HWNLFGLEINQVGYQGQVLPMLVATYILATIEKGLRKVIPTVLDNLLTPLLAILSTGF
                     ITFSFVGPLTRTLGYWLSDGLTWLYEFGGAIGGLIFGLLYAPIVITGMHHSFIAIETQ
                     LIADSSSTGGSFIFPIATMSNIAQGAAALAAFFIIKENKKLKGVASAAGVSALLGITE
                     PAMFGVNLKLRYPFIGAIVGSGIGSAYIAFFKVKAIALGTAGIPGFISISGQNNGWLH
                     YGIAMIIAFIVAFGVTYALSYRKKYRNIEA"
BASE COUNT      909 a    347 c    521 g    878 t
ORIGIN
```

```
  Query Match             13.8%;  Score 210.6;  DB 1;  Length 2655;
  Best Local Similarity   56.2%;  Pred. No. 1.9e-46;
  Matches  485;  Conservative    0;  Mismatches  359;  Indels   19;  Gaps    4;

Qy      5 tggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaagcgtt 64
          | ||   ||| || ||  | || |||||  |  |  ||||||| |  ||  | || || |
Db   1537 TTGCTAATGCCCCATTTACATTATTGCCTATACTAATTGGTTTTAGTGCTGCAAAACGAT 1596

Qy     65 tcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgagcttg 124
          | || |||||||  | ||| || || ||   |||| |||    ||||    ||    |||
Db   1597 TTGGAGGCAATGCATACCTCGGAGCGGCAC-TTGGAATGATATTGGTTCATCCAGAATTG 1655

Qy    125 gtgaacggctacgacgtggccgccaccatggctgcgggcgaa---atgccaatgtggtcc 181
           |||  |  || ||     |     |  | |  |||||  ||   || ||    |||  |
Db   1656 ATGAGTGCTTATGATTATCCTAAAGCATTAGAAGCGGGAAAAGAGATTCCTCATTGGAAC 1715

Qy    182 ctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggtg 241
          | || ||| | ||  ||   ||||  || ||||| ||    ||  | ||  |||| ||
Db   1716 TTATTCGGTCTCGAAATTAATCAAGTAGGGTACCAAGGTCAAGTCTTACCAATGCTAGTA 1775

Qy    242 gtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgca 301
          | | | |  ||||| ||||| || || ||     ||     ||     |            |
Db   1776 GCTACATATATTCTTGCAACAATTGAAAAAGGACTAAGGAAAGTTATACCAACAGTATTA 1835

Qy    302 gacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgcc 361
          ||     |  | || ||  |  |  |  |  |    || || ||  | |||||
Db   1836 GATAATTTATTAACACCTTTATTAGCAATATTATCAACTGGCTTTATAACATTTTCATTT 1895

Qy    362 attggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggactttat 421
          ||||  |||  ||  ||      | ||  ||  |  |  || |||||  ||  | |  || |||
Db   1896 GTTGGTCCATTAACACGTACTTTAGGGTATTGGTTATCAGACGGTTTAACGTGGCTATAT 1955

Qy    422 gatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcatc 481
          || || ||||||  |  |||| || ||  | || ||| || | ||  | ||||||||  ||
Db   1956 GAATTTGGTGGTGCTATCGGTGGGCTTATATTTGGTTTGTTATATGCGCCAATCGTTATT 2015

Qy    482 actggtctgcaccagtccttcccgccaattgagctggagctgtttaacca---------- 531
          || ||| |||| ||    ||     | |||||   | | ||  |     |
Db   2016 ACAGGTATGCATCATAGTTTTATTGCTATTGAAACGCAACTTATCGCTGACAGTTCATCA 2075

Qy    532 --gggtggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggca 589
            |||||| || || || ||| | |  ||| | ||| ||||||| || || || |||||
Db   2076 ACGGGTGGTTCATTTATATTCCCTATTGCAACAATGTCTAATATTGCACAAGGAGCGGCT 2135

Qy    590 tgtttggcagtgttcttcctggcga---agagtgaaaagctcaagggccttgcaggtgct 646
            | |||| ||||||  |    |   | | | ||||  | || ||| | ||   ||||
Db   2136 GCGCTTGCAGCGTTCTTTATTATAAAAGAAAATAAAAAATTAAAAGGCGTAGCCTCTGCT 2195

Qy    647 tcaggtgtctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgc 706
          |||||| || ||| |  | ||||||||| || || || || || |||||||||  |
Db   2196 GCAGGTGTATCAGCTCTATTAGGTATTACAGAACCAGCAATGTTTGGTGTGAACTTGAAG 2255

Qy    707 ctgcgctggccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgca 766
          ||  | |  || ||  |    |  ||||      ||| ||| |||   || |  || |||
Db   2256 CTAAGATATCCATTTATAGGGGCAATCGTAGGTTCAGGTATTGGTTCTGCCTATATAGCA 2315

Qy    767 ctctttaatatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctatt 826
          |||||||  | || ||  |||| || ||  |||||||| |    ||  || | ||||||
Db   2316 TTCTTTAAGGTTAAAGCTATTGCATTAGGTACTGCAGGTATACCAGGATTTATATCTATT 2375

Qy    827 gatgctccagatatggtcatgtt 849
            || || | |||  |    |||
Db   2376 AGTGGTCAAAATAATGGTTGGTT 2398


RESULT  14
BSTREAPR
LOCUS       BSTREAPR               4158 bp    DNA     linear   BCT 21-NOV-1996
DEFINITION  B.subtilis treA, treP and treR genes.
ACCESSION   Z54245
VERSION     Z54245.1  GI:1000450
KEYWORDS    enzyme II-tre; phospho-alpha-(1,1)-glucosidase; phosphotrehalase;
            repressor; treA gene; trehalose operon; treP gene; treR gene.
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 4158)
  AUTHORS   Schock,F. and Dahl,M.K.
  TITLE     Analysis of DNA flanking the treA gene of Bacillus subtilis reveals
            genes encoding a putative specific enzyme IITre and a potential
            regulator of the trehalose operon
  JOURNAL   Gene 175 (1-2), 59-63 (1996)
  MEDLINE   97074649
REFERENCE   2  (bases 1 to 2543)
  AUTHORS   Helfert,C., Gotsche,S. and Dahl,M.K.
  TITLE     Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed
            by a phospho-alpha-(1-1)-glucosidase encoded by the treA gene
  JOURNAL   Mol. Microbiol. 16 (1), 111-120 (1995)
  MEDLINE   95379486
REFERENCE   3  (bases 1 to 4158)
  AUTHORS   Dahl,M.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1995) Dahl M. K., University of
            Erlangen-Nuernberg, Lehstuhl fuer Mikrobiologie, Staudtstr. 5,
            Erlangen, Fed. Rep. of Germany, 91058
FEATURES             Location/Qualifiers
     source          1. .4158
                     /organism="Bacillus subtilis"
                     /strain="168 Marburg"
                     /isolate="1A1"
                     /db_xref="taxon:1423"
                     /clone="clone= pCH1"
                     /sex="-"
                     /clone_lib="genomic DNA"
     gene            167. .1579
                     /gene="treP"
     CDS             167. .1579
                     /gene="treP"
                     /function="enzyme II tre, domains B and C"
                     /note="position 1136 corresponds to position 1 of x80203"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="TreP"
                     /protein_id="CAA91014.1"
                     /db_xref="GI:1000451"
                     /db_xref="SWISS-PROT:P39794"
                     /translation="MGELNKSARQIVEAVGGAENIAAATHCVTRLRFALIDESKVDQE
                     MLDQIDVVKGSFSTNGQFQVVIGQGTVNKVYAELVKETGIGESTKDEVKKASEKNMNP
                     LQRAVKTLADIFIPILPAIVTAGLLMGINNILTAEGISFSTKSIVQVYPQWADLANMI
                     NLIAGTAFTFLPALIGWSAVKRFGGNPLLGIVLGVMLVHPDLLNAWGYGAAEQSGEIP
                     VWNLFGLEVQKVGYQGQVLPILLASYMLAKIEVFLTKRTPEGIQLLVVAPITLLLTGF
                     ASFIIIGPITFAIGNVLTSGLISVFGSFAALGGLLYGGFYSALVITGMHHTFLAVDLQ
                     LIGSKLGGTFLWPMLALSNIAQGSAALALMFIVKDEKQKGLSLTSGISAYLGITEPAI
                     FGVNLRYRFPFIIAMVSSGLAGMYISSQGVLASSVGVGGVPGIFSIMSQYWGAFAIGM
                     AIVLIVPFAGTYAYGRFKHK"
     gene            1650. .3317
                     /gene="treA"
     CDS             1650. .3317
                     /gene="treA"
                     /standard_name="phosphotrehalase"
                     /function="phospho-alpha-(1,1)-glucosidase"
                     /citation=[2]
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="TreA"
                     /protein_id="CAA91015.1"
                     /db_xref="GI:1000452"
                     /db_xref="SWISS-PROT:P39795"
                     /translation="MKTEQTPWWKKAVVYQIYPKSFNDTTGNGVGDLNGIIEKLDYLK
                     TLQVDVLWLTPIYDSPQHDNGYDIRDYYSIYPEYGTMEDFERLVSEAHKRDLKVVMDL
                     VVNHTSTEHKWLRGDLFNRQPYRDFYIWKKPQENGSVPTNWESKFGGSAWELDEASGQ
```

```
                     YYLHLFDVTQADLNWENEEVRKHVYDMMHFWFEKEIDGFRLDVINLISKDQRFPNAEE
                     GDGRSFYTDGPRVHEFLHEMNEKVFSHYDSMTVGEMSSTTVDHCIRYTNPDNKELDMT
                     FSFHHLKVDYLTAEKWALAPFDFLKLKEILSDWQTGMHAGGWNALFWCNHDQPRVVS
                     RYGDDGAYRVKSAKMLATAIHMMQGTPYIYQGEELGMTNPKFTDISSYRDAESLNMYH
                     AFKEKGMADQDITAILQAKSRDNSRTPVQWDATENGGFTTGTPWIPVAGNYREINAEA
                     ALRDQNSVFYHYQKLIQIRKMYDIVTEGTYEIIAKDDPNIFAYLRHGSNEKLLVINNF
                     YGTEAAFTLPDSLAPDEWKAEVLLTNDEAREGLQNMTVAHMSPLFIV"
     gene            3352. .4068
                     /gene="treR"
     CDS             3352. .4068
                     /gene="treR"
                     /function="repressor of the trehalose operon"
                     /note="position 3684 corresponds to position 2543 of
                     x80203"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="TreR"
                     /protein_id="CAA91016.1"
                     /db_xref="GI:1000453"
                     /db_xref="SWISS-PROT:P39796"
                     /translation="MKVNKFITIYKDIAQQIEGGRWKAEEILPSEHELTAQYGTSRET
                     VRKALHMLAQNGYIQKIRGKGSVVLNREKMQFPVSGLVSFKELAQTLGKETKTTVHKF
                     GLEPPSELIQKQLRANLDDDIWEVIRSRKIDGEHVILDKDYFFRKHVPHLTKEICENS
                     IYEYIEGELGLSISYAQKEIVAEPCTDEDRELLDLRGYDHMVVVRNYVFLEDTSLFQY
                     TESRHRLDKFRFVDFARRGK"
BASE COUNT     1137 a    838 c   1088 g   1095 t
ORIGIN


  Query Match             13.5%;  Score 205.4;  DB 1;  Length 4158;
  Best Local Similarity   53.1%;  Pred. No. 5.4e-45;
  Matches  485;  Conservative    0;  Mismatches  421;  Indels    7;  Gaps    2;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          || || |||    || | ||  |||| ||||| | |||  | || |   | ||   ||||
Db    650 CTGATTGCAGGCACGGCCTTTACGTTTTTGCCTGCGTTAATCGGATGGTCTGCCGTCAAG 709

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          || ||||||||||||  | | || |||   | |  | ||  |   | | |||   ||
Db    710 CGGTTCGGCGGCAATCCGCTTCTCGGCATTGTGC-TCGGGGTTATGCTCGTGCATCCTGA 768

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
           ||| |||| |  |  |    | |||  |          || || || ||  | |||
Db    769 TTTGCTGAATGCGTGGGGATATGGCGCGGCAGAACAAAGCGGAGAGATCCCGGTATGGAA 828

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          || || ||  |||| ||     ||| ||| || ||||||   ||||| ||  |  || |
Db    829 TCTATTCGGCCTAGAGGTGCAGAAAGTCGGCTATCAGGGCCAGGTGCTCCCAATTTTGCT 888

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          | |||||  || ||||| | ||| |||   ||  |    || ||       |
Db    889 TGCTTCTTATATGCTGGCGAAGATTGAGGTATTTTTAACAAAACGCACACCTGAGGGCAT 948

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          | |  | ||  |     ||    | ||| | ||  | || |||||     | || |||
Db    949 ACAGCTGCTCGTTGTCGCACCAATTACGCTTCTCTTAACAGGATTTGCCTCTTTTATCAT 1008

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          || |||||    |       | | ||  |||| ||| |    ||  ||        | |
Db   1009 TATCGGCCCGATTACATTTGCGATCGGTAATGTACTGACTTCAGGGCTTATCTCGGTGTT 1068

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          || ||    ||   || | ||||||||| | |  ||     |||||||| |  |||| ||
Db   1069 TGGTTCTTTTGCCGCACTGGGCGGTCTGTTATATGGCGGTTTCTACTCAGCGCTCGTGAT 1128

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagct------gtttaaccaggg 534
          || ||| |||| ||  | || |   || | || ||| ||||       |  || |  ||
Db   1129 TACCGGTATGCATCACACGTTTCTTGCAGTCGACCTGCAGCTCATCGGTTCAAAGCTCGG 1188

Qy    535 tggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgttt 594
          ||  |||| | |   | | |    |  |  |||||||||| || ||| ||||   | |

Db   1189 CGGCACCTTTTTATGGCCGATGCTGGCGCTTTCTAATATCGCACAAGGTTCGGCGGCTCT 1248

Qy     595 ggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgt 654
           |||| || | ||  | | |||| ||| || |  || |||||  |     | |||||  |
Db    1249 GGCATTGATGTTTATCGTGAAGGATGAGAAACAAAAAGGCCTCTCCCTCACATCAGGGAT 1308

Qy     655 ctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctg 714
           || |||  |||||| ||||| |||||||| ||||| || ||||| || ||     | |
Db    1309 TTCTGCTTATCTTGGCATTACAGAGCCTGCCATCTTTGGGGTGAATCTGCGATACAGATT 1368

Qy     715 gccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaa 774
           |||||  |||| |  || |  | | | |   | |  ||    |  ||| |
Db    1369 CCCGTTTATCATTGCAATGGTCAGCTCTGGCCTGGCCGGAATGTATATTTCTTCTCAAGG 1428

Qy     775 tatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctcc 834
           | |   ||||    |  | || |  |  ||  |   |||  || | || ||       |
Db    1429 TGTGCTGGCAAGCTCCGTCGGTGTCGGCGGCGTGCCGGGGATTTTCTCAATCATGAGCCA 1488

Qy     835 agatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagc 894
           |   || |  |||     |  |    | | ||   ||  || |  | || | || | |
Db    1489 GTACTGGGGCGCGTTTGCAATCGGGATGGCTATTGTATTGATCGTGCCGTTTGCCGGAAC 1548

Qy     895 gattgcttatggc 907
             || ||||||
Db    1549 ATACGCGTATGGC 1561


RESULT  15
D83967/c
LOCUS       D83967                 22197 bp    DNA     linear   BCT 20-NOV-1997
DEFINITION  Bacillus subtilis genomic DNA, 74 degree region.
ACCESSION   D83967
VERSION     D83967.1  GI:2626809
KEYWORDS    TreP; TreC; TreR; YfkP; YfkO; YfkN; YfkM; YfkL; YfkK; YfkJ; YfkI;
            YfkH; YfkG; YfkF; YfkE; YfkD; YfkC; YfkB; YfkA; YfjT; YfjS; YfjR.
SOURCE      Bacillus subtilis (strain:AC327) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 22197)
  AUTHORS   Sekiguchi,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1996) Junichi Sekiguchi, Textile Science and
            Technology, Shinshu University, Department of Applied Biology;
            3-15-1 Tokida, Ueda-shi, Nagano 386, Japan
            (E-mail:jsekigu@giptc.shinshu-u.ac.jp, Tel:0268-21-5344,
            Fax:0268-21-5331)
REFERENCE   2  (bases 1 to 22197)
  AUTHORS   Sekiguchi,J., Yamamoto,H., Uchiyama,S. and Fajar,A.
  TITLE     Nucleotide sequence analysis of B. subtilis cromosome in 74 degree
            region
  JOURNAL   Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1. .22197
                     /organism="Bacillus subtilis"
                     /strain="AC327"
                     /db_xref="taxon:1423"
     RBS             47. .64
     CDS             62. .847
                     /codon_start=1
                     /transl_table=11
                     /product="YfjR"
                     /protein_id="BAA23388.1"
                     /db_xref="GI:2626810"
                     /translation="MTVYNRTKQKTEDLVTEGAQAADTPRLAAKSADIVITMLADDDS
                     VSTVTFGEDGLLEGLAENGIHISMSTISVEFSEKLAAAHAEKGQFFLAAPVLGRPDAA
                     AKAALRIITAGPAEAKQAAKPLLDSLSQQIFDVGEESKTANAAKISINFLLVSMLEAL
                     SESFLMMEKYGLEQKQFLEIASALFGSPVYQNYGTIMAEQKFEPAGFKMSLGLKDTNL
                     ALAAAKRVSANLPLAELAKSHFESGIEKGFGDLDWAALIKCIK"
     terminator      complement(855. .874)
     CDS             complement(885. .1676)
                     /codon_start=1
                     /transl_table=11
```

/product="YfjS"
/protein_id="BAA23389.1"
/db_xref="GI:2626811"
/translation="MKWMCSICCAAVLLAGGAAQAEAVPNEPINWGFKRSVNHQPPDA
GKQLNSLIEKYDAFYLGNTKEKTIYLTFDNGYENGYTPKVLDVLKKHRVTGTFFVTGH
FVKDQPQLIKRMSDEGHIIGNHSFHHPDLTTKTADQIQDELDSVNEEVYKITGKQDNL
YLRPPRGVFSEYVLKETKRLGYQTVFWSVAFVDWKINNQKGKKYAYDHMIKQAHPGAI
YLLHTVSRDNAEALDDAITDLKKQGYTFKSIDDLMFEKEMRLPSL"
RBS complement(1674. .1691)
terminator complement(1754. .1778)
CDS complement(1777. .1962)
/codon_start=1
/transl_table=11
/product="YfjT"
/protein_id="BAA23390.1"
/db_xref="GI:2626812"
/translation="MGNAVNNKDQQIDYLKNRLDMFMNVIDSLDPEATDVEDIDRLIS
MLDDLEAKYERFKKDWE"
RBS complement(1960. .1976)
RBS 2090. .2109
CDS 2107. .2571
/codon_start=1
/transl_table=11
/product="YfkA"
/protein_id="BAA23391.1"
/db_xref="GI:2626813"
/translation="MTQNTKLRPITPEFDPWEAYMDVDQYGDMQLTNVEFTTTTLCNM
RCEHCAVGYTLQPKDPNALPIDLLLKRLEEIPRLRSISITGGEPMLSLKSVKEYVVPL
LKYAHERGVRTQINSNLTLDIERYEWIIPYLDVLHISHNWGTVEDFAEIGFA"
terminator complement(2666. .2701)
RBS 2750. .2768
CDS 2766. .3227
/codon_start=1
/transl_table=11
/product="YfkB"
/protein_id="BAA23392.1"
/db_xref="GI:2626814"
/translation="MYPSDFASALESLSLKEMRQAIHRLLDIRDENTWMLFGTLPFYA
CSPDEDDQKLLRRLRAAKNVTVRNDPDGRSRLNVNIFDGNIIVTDFGDTPPLGNIQTD
SLPSAYEKWRETKLAKELSCHCPNVRCLGPNVLVKNSYYQDTDFTKRQARS"
RBS 3217. .3230
CDS 3228. .4070
/codon_start=1
/transl_table=11
/product="YfkC"
/protein_id="BAA23393.1"
/db_xref="GI:2626815"
/translation="MRMKETLTEIFQNKIVDILLVAVILWIGVFIINRLVQLFFKRTD
FIEEKKEKTIESLVRSVTQYTATIGFIFYVISLFVHDFGKILAGAGVAGIVIGFGAQS
LIKDVLAGVFLIYERQLHKGDYVTVNNLFNGTVEEIGLRSLQIREWSGKLLTISNGEV
RQIENYNIDFMRITESFLISFKEDPDRVYSVLEEACDMLNEELRDSLKRDEFGNPTEP
FQIHGITALNKINRGVEFTVKGMVKDDDYFSASLAVRRVLVRQLYQNNVQMLEEAVRI
ERTQ"
terminator complement(4074. .4106)
CDS complement(4109. .4903)
/codon_start=1
/transl_table=11
/product="YfkD"
/protein_id="BAA23394.1"
/db_xref="GI:2626816"
/translation="MMKKLFHSTLIVLLFFSFFGVQPIHAKKQFKVPNSVASISKENT
YPNASQDQPMLQPSKLAKELLDHSEVKIENPHLIKMLNESNISGTPLAVGYRATIFLG
KWALGYESNETVANWEYKKINTNRADNRGGKETAEMHYAQEQQYRVKGGLTAKVPNAE
DVKSMMMQKAMKKTNLPLAFETVIGAGTKRDQIYKVAPKKIGYLHAYAPAVNEKGKVT
YGEVYLVLKGNKRKLVVKNVTSQGIGAWIPVQDHVTFGFQLSSLPR"
RBS complement(4901. .4913)
CDS complement(4974. .6029)
/codon_start=1
/transl_table=11
/product="YfkE"
/protein_id="BAA23395.1"
/db_xref="GI:2626817"
/translation="MNRIFFILVAAGVPLSVIGSLMHWPSAVLFAVYCVTIIALASYM
GRATESLSIIAGPRIGGLLNATFGNAVELIISLFALKEGLTGIVLASLTGSVLGNLLL
VAGLSFFVGGLKYKRQEFNIHDARHNSGLLIFAIIVAFVIPEVFSVGMGNASKLNLSI
GISIIMILLYVAALYFKLVTHRGVYQPNNAAQTEEEEEPEWSGKVATIVLFAATIVVA
YISENLVHTFHSVAEQFGWSELFIGVIIVAIVGNAAEHASAIIMAFKNKMDIAVEIAV
GSTLQIAMFVAPVLVICSIFFPTSMPLVFTLPELVAMVSAVLLMIAISNDGDSNWFEG
ATLLAAYVIMAIGFFLL"
RBS complement(6027. .6046)
RBS 6180. .6199
CDS 6197. .7372
/codon_start=1
/transl_table=11
/product="YfkF"
/protein_id="BAA23396.1"
/db_xref="GI:2626818"
/translation="MSRFHFFILVLLVSISGFSQGMLLPVISIIFETNGESAAINGLH
ATGLYIGVLLASPFMEAPLRKLGFKPLIVMGGSIVILSLFGFIWLQSVWVWFLLRLFI
GIGDHMLHFSTQTWVTSMSSKQNRGRNLSIYGLSFGLGFAAGPFMVPLVKLSPSLPFI
VSGCISLFAWLFVFFLQNAYPETSPHETKSDNSFRRFYQAMLFGWVAFMPTFGYGFLE
TALNGSFPVYALRLGISVDAVAIILPAFAIGSIIFQFPLGILSDKYGRRNVLLVILLT
GALCFFIAGVFPSPYVIGGCFFIAGMAVGSTFTLGISYMTDLLPPHLLPAGNLLCGIT
FSLGSILGPVAGGWYMQTFESANLFYFITLTLSSVWLALVLGKPKSWSPAETYSSSS"
terminator complement(7376. .7408)
RBS 7396. .7411
CDS 7409. .7546
/codon_start=1
/transl_table=11
/product="YfkG"
/protein_id="BAA23397.1"
/db_xref="GI:2626819"
/translation="MLMVAVLSVGASACFMARFAFRVLFVMASAFFRMLGHHDIPPTS
F"
CDS complement(7571. .8398)
/codon_start=1
/transl_table=11
/product="YfkH"
/protein_id="BAA23398.1"
/db_xref="GI:2626820"
/translation="MSFLKELFSRYTLHEGQSKSAELAYFFLLSLFPFLIFMLTLTAY
LPLSTDDVLGVIEQYAPASAMSLVESITHQTLNNRNGGLLSFGIIAALWSASNGMNAI
VRSLNHAYDVEENRSFIIVRLTSIFLTIAMVFTILVALLLPVFGREIGRLASDFVGAS
DLFLSVWAAIRWGVSPLVLLIVFSALYVIAPNKKLSLRFVMPGAVFATIGWIIVSTLF
SFYVSTFANYSATYGSIGGIIVLMIWFYLSGILIILGGEINALLHKRKKLPDENPYH"
RBS complement(8396. .8415)
CDS complement(8422. .8742)
/codon_start=1
/transl_table=11
/product="YfkI"
/protein_id="BAA23399.1"
/db_xref="GI:2626821"
/translation="MNNERLMLKGIFLGAAAGAALSLLHKPTRQACGMRWLTCKHKLS
LYKSNPELLKNTVITKVDEAKKLARTLSKEVDFVNQQVKELKKTTPQVMELVQETKEH
FSKK"
RBS complement(8740. .8756)
CDS complement(8760. .9230)
/codon_start=1
/transl_table=11
/product="YfkJ"
/protein_id="BAA23400.1"
/db_xref="GI:2626822"
/translation="MISVLFVCLGNICRSPMAEAIFRDLAAKKGLEGKIKADSAGIGG
WHIGNPPHEGTQEILRREGISFDGMLARQVSEQDLDDFDYIIAMDAENIGSLRSMAGF
KNTSHIKRLLDYVEDSDLADVPDPYYTGNFEEVCQLIKTGCEQLLASIQKEKQL"
RBS complement(9228. .9247)
RBS 9418. .9435
CDS 9433. .9648
/codon_start=1
/transl_table=11
/product="YfkK"
/protein_id="BAA23401.1"
/db_xref="GI:2626823"
/translation="MSSPNTETLTQMIEEISQKLNMLNVGVIKAEDFSDEKIEDLTYL
HRMVMKKESFSPSEMQAIAQELASLRK"
RBS 9727. .9743

```
     CDS             9741. .10931
                     /codon_start=1
                     /transl_table=11
                     /product="YfkL"
                     /protein_id="BAA23402.1"
                     /db_xref="GI:2626824"
                     /translation="MQTTSIKTASGMYINYFFLGMVNIILASNMSSLTKQWNTDPTGI
                     SYVIAAIGFGKLLTYGISGVLSDKIGRKPLVVASAGIMAVFLVGIPLSPSYELAFVFA
                     LLAGVANSAMDAGTYPALTELFPSASGSANVLVKAFMSVGAALLPLLITFLADHSMFY
                     GFAFYLPAAVYLLNIIYLSTLSFPKKHKKPTNSGQQESPVFLSEPVFQKEGTALIIIG

  Query Match           13.2%;  Score 201.8;  DB 1;  Length 22197;
  Best Local Similarity  52.9%;  Pred. No. 6.7e-44;
  Matches  482;  Conservative    0;  Mismatches  422;  Indels    7;  Gaps    2;

Qy      1 ctcatggcatctgcgccgttcgcgttcttgccagtgttggttggtttcaccgcaaccaag 60
          || || |||    || | ||  |||| ||||| | |||  | || |   | ||   ||||
Db  20384 CTGATTGCAGGCACGGCCTTTACGTTTTTGCCTGCGTTAATCGGATGGTCTGCCGTCAAG 20325

Qy     61 cgtttcggcggcaatgagttcctgggcgccgcgtattggtatggcgatggtgttcccgag 120
          || |||||||||||  | | || |||   | |  | ||  |   | | |||   ||
Db  20324 CGGTTCGGCGGCAATCCGCTTCTCGGCATTGTGC-TCGGGGTTATGCTCGTGCATCCTGA 20266

Qy    121 cttggtgaacggctacgacgtggccgccaccatggctgcgggcgaaatgccaatgtggtc 180
           ||| |||| |  |  |      | |||  |           || || || ||  | |||
Db  20265 TTTGCTGAATGCGTGGGGATATGGCGCGGCAGAACAAAGCGGAGAGATCCCGGTATGGAA 20206

Qy    181 cctgtttggtttagatgttgcccaagccggttaccagggcaccgtgcttcctgtgctggt 240
          || || ||  |||| ||     ||| ||| || ||||||   |||| ||  |  || |
Db  20205 TCTATTCGGCCTAGAGGTGCAGAAAGTCGGCTATCAGGGCCAGGTGCTCCCAATTTTGCT 20146

Qy    241 ggtttcttggattctggcaacgatcgagaagttcctgcacaagcgactcaagggcactgc 300
          | ||||  || |||| | ||| |||   ||  |    || ||        |
Db  20145 TGCTTCTTATATGCTGGCGAAGATTGAGGTATTTTTAACAAAACGCACACCTGAGGGCAT 20086

Qy    301 agacttcctgatcactccagtgctgacgttgctgctcaccggattccttacattcatcgc 360
          | |  | ||  |     ||    | ||| | ||  | || |||||     | || |||
Db  20085 ACAGCTGCTCGTTGTCGCACCAATTACGCTTCTCTTAACAGGATTTGCCTCTTTTATCAT 20026

Qy    361 cattggcccagcaatgcgctgggtgggcgatgtgctggcacacggtctacagggacttta 420
          || |||||    |       | | ||  |||| ||| |    || ||        | |
Db  20025 TATCGGCCCGATTACATTTGCGATCGGTAATGTACTGACTTCAGGGCTTATCTCGGTGTT 19966

Qy    421 tgatttcggtggtccagtcggcggtctgctcttcggtctggtctactcaccaatcgtcat 480
          || ||    ||   || | |||||||||| | |  ||     |||||||| |  |||| ||
Db  19965 TGGTTCTTTTGCCGCACTGGGCGGTCTGTTATATGGCGGTTTCTACTCAGCGCTCGTGAT 19906

Qy    481 cactggtctgcaccagtccttcccgccaattgagctggagct------gtttaaccaggg 534
          || ||| |||| ||  | || |   || | || ||| ||||       |  || |  ||
Db  19905 TACCGGTATGCATCACACGTTTCTTGCAGTCGACCTGCAGCTCATCGGTTCAAAGCTCGG 19846

Qy    535 tggatccttcatcttcgcaacggcatctatggctaatatcgcccagggtgcggcatgttt 594
          ||  |||| | |   | | |    |  |  |||||||||| || ||| ||||   | |
Db  19845 CGGCACCTTTTTATGGCCGATGCTGGCGCTTTCTAATATCGCACAAGGTTCGGCGGCTCT 19786

Qy    595 ggcagtgttcttcctggcgaagagtgaaaagctcaagggccttgcaggtgcttcaggtgt 654
          |||  || | ||  | | ||||  ||| || |  || |||||  |     | |||||  |
Db  19785 GGCCATGATGTTTATCGTGAAGGATGAGAAACAAAAAGGCCTCTCCCTCACATCAGGGAT 19726

Qy    655 ctccgctgttcttggtattacggagcctgcgatcttcggtgtgaaccttcgcctgcgctg 714
          || |||  |||||| ||||| |||||||| ||||| || ||||| || ||      | |
Db  19725 TTCTGCTTATCTTGGCATTACAGAGCCTGCCATCTTTGGGGTGAATCTGCGATACAGATT 19666

Qy    715 gccgttcttcatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaa 774
          |||||  |||| |  || |  | | | |   | |  ||    |  ||| |
Db  19665 CCCGTTTATCATTGCAATGGTCAGCTCTGGCCTGGCCGGAATGTATATTTCTTCTCAAGG 19606

Qy    775 tatcaaggcagttgcgttgggcgctgcaggtttcttgggtgttgtttctattgatgctcc 834
          | |   ||||    |  | || |  |  ||  |   |||  || | || ||       |
Db  19605 TGTGCTGGCAAGCTCCGTCGGTGTCGGCGGCGTGCCGGGGATTTTCTCAATCATGAGCCA 19546

Qy    835 agatatggtcatgttcttggtgtgtgcagttgttaccttcttcatcgcattcggcgcagc 894
            |  || |  |||     |  |    | | ||   ||  || |  | || | || | |
```

```
Db  19545 GTACTGGGGCGCGTTTGCAATCGGGATGGCTATTGTATTGATCGTGCCGTTTGCCGGAAC 19486

Qy    895 gattgcttatg 905
          .     || ||||
Db  19485 ATACGCGTATG 19475


Search completed: September 25, 2002, 07:30:47
Job time: 5098 sec
```